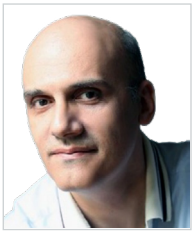
A Comprehensive Analysis of the Importance of Programming in Modern Data Journalism

Un análisis exhaustivo de la importancia de la programación en el periodismo de datos moderno



Andreas Veglis is a professor of media technology, and head of the Media Informatics Lab at the School of Journalism and Mass Communication at Aristotle University of Thessaloniki. He is serving and has served as an editor, member of scientific boards, and reviewer in various academic journals. Prof. Veglis has more than 200 peer-reviewed papers on media technology and journalism. Specifically, he is the author or co-author of 12 books and 44 book chapters, he has published more than 115 papers in scientific journals, and he has presented 154 papers at international and national Conferences. Prof Veglis has been involved in 50 national and international research projects. His research interests include information technology in journalism, new media, algorithmic journalism, in journalism, digital security, data journalism, big data, and content verification.
Aristotle University of Thessaloniki, Greece
veglis@jour.auth.gr
ORCID: 0000-0002-0286-2304

ISSN: 1696-019X / e-ISSN: 2386-3978

Received: 31/12/2024 - Accepted: 17/02/2025 - Early access: 26/02/2026 - Published:

Recibido: 31/12/2024 - Aceptado: 17/02/2025 - En edición: 26/02/2026 - Publicado

Abstract:

In the last 20 years, the field of data journalism has seen significant development. Data journalism is a specialization based on finding news stories in data. Today, there is a trend in data journalism to use programming techniques (usually Python or R). This trend comes mainly from prominent journalism organizations engaged in big data journalism projects based on data analysis. The main thrust of this study is to examine the importance of programming in data journalism today. A mixed methodology is employed, including a literature review and case study analysis. The findings indicate that although coding is not necessary in data journalism, it may be viewed as a necessity when used for complex projects that include large datasets and specific types of visualization. The study also discusses hybrid approaches and

Resumen:

En los últimos 20 años, se ha experimentado un desarrollo significativo en el campo del periodismo de datos, que es una especialización basada en la búsqueda de noticias en los datos. Hoy en día, existe una tendencia en el periodismo de datos a emplear técnicas de programación (normalmente Python o R). Esta tendencia proviene principalmente de importantes organizaciones periodísticas que participan en proyectos de periodismo de los macrodatos o big data basados en el análisis de datos. El objetivo principal de este estudio es examinar la importancia de la programación en el periodismo de datos actual. Se emplea una metodología mixta, que incluye una revisión bibliográfica y un análisis de estudios de caso. Los hallazgos indican que, si bien la codificación no es necesaria en el periodismo de datos, puede considerarse una necesidad

Cómo citar este artículo:
Veglis, A. (2026). Un análisis exhaustivo de la importancia de la programación en el periodismo de datos moderno. *Doxa Comunicación*, 43.
https://doi.org/10.31921/doxacom.2878

specific tools that aim to bridge the gap between tool-based and code-based approaches in practicing data journalism. The results indicate that programming should be viewed as a complementary skill that should be considered by data journalists.

Keywords:

Data journalism, programming, coding, Python, R, case study methodology.

cuando se utiliza para proyectos complejos que incluyen grandes conjuntos de datos y tipos específicos de visualización. El estudio también analiza enfoques híbridos y herramientas específicas que tienen como objetivo cerrar la brecha entre los enfoques basados en herramientas y los basados en código en la práctica del periodismo de datos. Los resultados indican que la programación debería considerarse una habilidad complementaria que los periodistas de datos deberían considerar.

Palabras clave:

Periodismo de datos, programación, codificación, Python, R, metodología de estudio de caso.

ISSN: 1696-019X / e-ISSN: 2386-3978

1. Introduction

Since the conception of journalism, its relationship with technology has been symbiotic, with each shaping and influencing the other (Siapera and Veglis, 2012). Recently, this relationship has taken a transformative turn with the move towards datafication. The concept of datafication was introduced by Mayer-Schönberger and Cukier (2013), and it was defined as the process of transforming large quantities of information into a scalable resource in order to be used as knowledge production and thus add economic value. The introduction of datafication in journalism can be seen as a significant shift because it fundamentally alters the process of news production, dissemination, and consumption. It is worth mentioning that journalism has already undergone a transformation caused by the introduction of big data technologies (Coddington, 2015). This has fueled the development of data journalism, a journalism specialty that employs computational methods to analyze large datasets and extract meaningful information (Gray, Bounegru, & Chambers, 2012; Veglis and Bratsas, 2021).

Data journalism has faced skepticism from journalists who are resistant to technological changes. However, with the rise of digitalization, data started to influence newsrooms through metrics, resource allocation, topic selection, and news format, and thus gained more attention in scientific publications (Angelou et al., 2020). Various terms, such as "data-driven journalism," "interactive journalism," and "big data journalism" has been employed to depict this new form of journalism (Ausserhofer et al., 2020). Thus, datafication has gradually become an important parameter in the production of journalistic content, with data journalism playing a significant role in this process (Porlezza, 2023).

The journalist's datafication of their toolkit has provided them with the ability to report complex issues and provide empirically grounded narratives (Lewis & Westlund, 2015). It has also democratized the field, allowing for greater public engagement through interactive visualizations and open data platforms (Young et al., 2018). This transformation has resulted in the redefinition of the journalist's role, which now includes tasks such as data analyst, visual designer, and interpreter of complex socio-political phenomena (Ausserhofer et al., 2020). The renovated journalism space is a new kind of journalism in the shape of altered journalistic practices through the growing incorporation of data and programming. Ruppert et al. (2017) argue that data has turned into a tool of power, thus, making such influence felt in the public and private sectors, even in the domain of journalism. They claim that the arrival of data journalism is part of the general reconfiguration of power structures, with data,

code, and algorithms as the most important actors. On the other hand, data journalism was practiced before the introduction of coding and was mainly focused on statistical analysis through computer-assisted reporting (CAR) (Parasie & Dagiral, 2013).

The necessity of data journalism coding is often debated within the media industry. Proponents on one side say that coding skills are important for journalists (and especially data journalists) in the current media landscape. They argue that coding enables journalists to collect, analyze, and present data in new and creative ways, thus enabling them to detect hidden patterns and provide a thorough explanation of complicated issues. In addition, they state that the usefulness of coding for storytelling and investigative journalism (Simon, 2021) has increased. Porlezza (2023) argues that coding is a prerequisite for the practice of data journalism. Data journalism involves the collection, analysis, and visualization of data to tell stories and disclose insights. To deal with information effectively, journalists usually utilize programming languages (such as Python and R) to clean data, conduct statistical analyses, and create data visualizations. Coding skills enable journalists to use big data, data processing automation, and interactive data visualizations. However, it must be said that the integration of coding with journalism practice also raises some concerns. In particular, critics insist that journalists must keep their focus on storytelling and information-gathering skills and leave the technical part to data experts and developers. The fear is that mastering coding may cause journalists' attention to be diverted from their main duties, and thus, journalistic integrity may be compromised (Cruz, n.d.; Hannaford, 2015). Heravi and Lorenz (2020) argue that the use of coding in data journalism is a matter of debate. Even if coding talents may be useful within the framework of data-driven stories, it should be noted that not all data journalists are obliged to be coding specialists. Porlezza (2023) does not exactly oppose that conclusion, but he states that the knowledge of coding may strengthen a journalist's capacity to deal with data properly. Coding is not considered a core component of data journalism but rather a tool that can meaningfully assist in the process (Heravi & Lorenz, 2020).

Finding the equilibrium point between the pros and cons of coding in journalism is crucial. Journalists endowed with programming skills can mine vivid data and attract audiences using various new techniques; however, it is important that journalistic standards and ethics are not compromised. Colman et al. (2018) explain the ethical issues that arise when coding is applied by journalists in their work.

Thus, the use of coding in journalism presents both opportunities and challenges. It is essential for journalists to critically evaluate the benefits and risks associated with coding, considering the specific context and goals of their reports. By leveraging coding skills responsibly and ethically, journalists can harness the power of technology to enhance their storytelling and deliver impactful journalism.

The main thrust of this study is to examine the significance of programming in data journalism practice today. Based on the previous discussion, two research questions were formulated. Specifically, the disagreement over whether programming is essential or supplementary in modern data journalism is the basis for the first research question:

RQ1: Is coding considered a necessity for practicing data journalism?

Additionally, the discussion in cases where programming is utilized points to the idea that the characteristics of a specific project likely determine the need for code-based approaches, thus helping us articulate the second research question:

RQ2: What are the features of a data journalism project that influence the decision to utilize coding for it?

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

1.1. Practicing data journalism

Simon Rogers is credited with first using the term data journalism in a post on the Guardian Insider Blog, as noted by Knight (2015). This concept encompasses a workflow that begins with data analysis, followed by filtering and visualizing the data to complement the narrative, as outlined by Lorenz (2010). It integrates aspects such as spreadsheets, graphics, data analysis, and major news stories (Rogers 2008) and is essentially about creating news graphics, incorporating design and interactivity elements (Bradshaw 2018; Lorenz 2010; Rogers 2008). Megan Knight (2015) defines data journalism as a narrative primarily based on numerical data or one that significantly involves data or visualization.

Veglis and Bratsas (2017a) offered a definition that more effectively encapsulates the importance of visualization and interactivity in data journalism. They describe it as a process of extracting valuable insights from data, crafting articles based on this information, and integrating visualizations (sometimes interactive) into these articles to aid reader comprehension or allow them to engage with data relevant to them. They also proposed the categorization of the data journalism process into six distinct stages: Data Compilation, Data Cleaning, Data Understanding, Data Validation, Data Visualization, and Article Writing (2017b).

Data Compilation: This initial phase of a data journalism project begins with either a question requiring data or a dataset that needs to be explored. Data compilation can occur in various ways, such as receiving data directly from an organization (often as open data), using advanced search techniques (mainly with the help of Google advanced search), searching with search engines that focus on datasets (for example, Data Search tool), web scraping (with dedicated scraping tools or through widely used applications that support web scraping), converting document formats to enable analysis (for example, extracting tables from PDF files and exporting them to CSV format), or gathering data through observation, surveys, online forms, or crowdsourcing (Veglis & Bratsas, 2017a).

Data Cleaning: Also known as data scrubbing, involves detecting and correcting erroneous or corrupted records within a dataset (Wu, 2013). It involves removing human errors and standardizing data formats for consistency with other data used by journalists (Veglis & Bratsas, 2017b). For this task, general-purpose applications such as Microsoft Excel and Google Sheets can be utilized, which include many features, some of which can be used for data cleaning directly or indirectly (Bauzon et al., 2021; Guerrero et al., 2019; Setiyanto & Setiawan, 2022).

Data Understanding: In this stage, journalists must decipher various codes in datasets representing categories, classifications, or locations, along with specialized terminology. Additional data are often required to make the existing data meaningful. Journalists need to be data literate and capable of understanding, articulating, and critically analyzing data (Veglis & Bratsas 2017a). For this purpose, it may be necessary to find additional data; therefore, the tools of the previous stage may be employed. In addition, tools for combining datasets may be useful. Such a process can be accomplished more easily with applications that support easy data combination, such as Tableau (Balaji, et al., 2021; Batt, et al., 2020; Loth, 2019). It is worth noting that dataset combination can be accomplished with the help of general-purpose spreadsheet applications such as Microsoft Excel or Google Sheets; however, it is not an easy process because these applications do not directly support such processes (Guerrero, Guerrero, & Rauscher, 2019).

Data Validation: This involves cross-checking original data and acquiring additional information from sources to enrich the data (Silverman, 2014; Veglis, 2013). It is important to recognize that datasets, like any source, have inherent biases and objectives. Journalists must investigate the origin, purpose, and collection methodology of a dataset (Bradshaw, 2018). This can be done by exploring the dataset's creation history, finding dataset references, or using other information sources related to the investigative subject (Silverman, 2014; Veglis & Bratsas, 2017a). Thus, general-purpose searching techniques, such as those employed in the data compilation stage, may be utilized.

Data Visualization: It is the visual representation of abstract information for analysis and communication (Cairo, 2012). Because statistical information is abstract, transforming it into a physical representation requires an understanding of visual perception and cognition. Effective data visualization follows design principles based on human perceptions (Card et al., 1999; Few, 2013). Several applications (desktop or online services) can be used to create visualizations (Protopsaltis et al., 2020). Most of them offer multiple templates for creating different types of visualizations. Typical examples include Infogram, Piktochart, DataWrapper, and Tableau.

1.2. Available solutions for adopting coding.

In the media industry, there are two alternative solutions for utilizing coding: the programming languages Python and R (and in some cases Ruby) (Parasie & Dagiral, 2013). Both offer robust capabilities for data processing and visualization. Python, which is characterized by its simplicity and readability, can be used for tasks such as web scraping, data analysis, and interactive visualization. It offers extensive libraries, such as Pandas and Matplotlib, that enable journalists to efficiently handle large datasets and present complex information in an accessible format. Python supports integration with web frameworks, a feature that supports the creation of interactive and online presentations (Lutz, 2001).

The R programming language is highly regarded for its statistical analysis and data visualization capabilities, making it an ideal choice for in-depth analysis of complex datasets. It offers many packages that can support the discovery of hidden patterns and trends in the data. R Markdown offers a platform for combining code, visualizations, and narratives to enhance storytelling in data journalism (Crawley, 2012).

Both Python and R are open-source, which fosters a collaborative environment, with a vast community of developers contributing to their development and offering support. This aspect is crucial for journalists, who often rely on shared resources and community support to learn and troubleshoot. Additionally, there is an abundance of educational resources and online tutorials for both programming languages, which facilitates their adoption by journalists (Appelgren & Nygren, 2014; Bradshaw, 2018; Bounegru et al., 2018; Hamilton, 2016; Knight, 2015).

Setting up Python or R language on a computer is not an easy process. Nevertheless, there are alternative solutions that utilize cloud computing, such as Google Colab, which supports writing and executing Python code through a browser (Tock, 2019). Specifically, Google Colab is a free, cloud-based Jupyter Notebook environment that enables writing and executing Python in a web browser with free access to GPUs and TPUs, making it suitable for machine learning and data science (Burke, 2023). Thus, journalists do not need to set up a personal computer to use Python.

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

Another recent development is the integration of Excel with Python. Microsoft combines Python's advanced computational abilities with Excel's spreadsheet features (Kinnestand, 2023). Users can directly employ Python in Excel cells by leveraging Python libraries for data manipulation and visualization[1]. This integration in Excel allows journalists to gradually migrate to Python coding while using the familiar Excel environment.

Another approach that is available for working with Python involves utilizing AI interfaces like ChatGPT, with the help of specific plugins[2]. Of course, it is worth noting that such solutions are still at an early stage of development, with plugins appearing and disappearing rapidly. With the help of specific plugins, ChatGPT can serve as an intuitive interface with its natural language processing capabilities for journalists who may not possess much programming knowledge and are not eager or have the time to learn programming. This could involve querying datasets, generating statistical summaries, or creating data visualizations using text commands. Thus, the utilization of Python for data journalism becomes easier, allowing journalists to focus more on the narrative and less on the technical aspects of data handling (Knight, 2015). The use of ChatGPT as an interface with Python can also be viewed as a democratization of data journalism, allowing a broader range of journalists to engage in data-driven storytelling. This approach aligns with the trend towards more accessible and collaborative forms of journalism (Hamilton, 2016). By lowering the barrier to entry for data journalism, AI plays a significant role in the evolution of the field, making it more inclusive and diverse than before.

1.3. Tool based vs code-based data journalism

Based on the previous discussion, a direct comparison between tool-based and code-based approaches to practicing data journalism is conducted. The comparison is based on the possibilities offered by each approach in facilitating the various stages of developing data journalism, as depicted by Veglis and Bratsas (2017b) and presented in Section 1.1. To facilitate such a comparison, Table 1 was constructed. Table 1 includes indicative tools (services and apps) that can be used in the five data journalism stages in the tool-based approach, along with the available capabilities that the two mentioned programming languages (Python and R) offer directly or with the help of specific libraries or packages.

1 https://www.microsoft.com/en-us/microsoft-365/python-in-excel

2 https://openai.com/index/chatgpt-plugins/

Table 1. Direct comparison of the possibilities offered by the tool-based and code-based approaches.

Data Journalism Stage	Tools utilized	Coding possibilities (Python / R)
Data Compilation	*Search engines*: Google, Bing, etc. *Data Search tool*: Google Data Search *Web scraping*: Microsoft Excel, Google Sheet. *PDF scraping*: Tabula	Conduct web searches in Python through the Google Custom Search API and in R with the help of GoogleSearchR package. Web scraping is supported by both Python (i.e BeautifulSup4), and R (i.e rvest).
Data Cleaning	*General purpose application*: Microsoft Excel, Google Sheet. *Dedicated cleaning tool*: OpenRefine	Data cleaning can be implemented by both Python and R since both programming languages the support data cleaning operations
Data Understanding	*Applications/tools that support combining data sets:* Tableau, Microsoft Excel, Google Sheet.	Python supports the combination of datasets, primarily using pandas library. The same is valid for R.
Data Validation	No additional application/tools are employed	Not applicable
Data Visualization	*Visualization tools/applications:* Infogram, Datawrapper, Tableau, Piktochart etc.	Both programming languages possess extensive capabilities in creating visualizations.

ISSN: 1696-019X / e-ISSN: 2386-3978

From Table 1, one can easily conclude that both approaches can support all stages of data journalism. Nevertheless, it is worth noting that in the code-based approach, all data journalism stages are supported through the same platform (i.e., Google Colab), whereas in the tool-based approach, no interoperability exists, and thus journalists have to manually export/save and import/load the data from one tool to another. This adds complexity to the process and creates opportunities for errors that may slow the production of data journalism articles.

2. Methodology

For this study, a mixed approach was selected, which included a literature review and case studies. The case Study methodology is a widely used research approach that can support the study of complex phenomena or processes. It is a qualitative research technique widely adopted in various disciplines, including social sciences, education and business. The case study methodology involves an in-depth exploration of one or more cases, utilizing various data collection techniques such as interviews, documents, and observations. It offers a holistic perspective and generates rich and detailed insights into the research topic (Kratochwill et al., 2013; Tellis, 1997).

Despite its advantages in providing comprehensive and in-depth insights and contributing to theory building and practical applications, as Yin (2014) suggests, the case study methodology has limitations. Baxter and Jack (2008) caution about the limited generalizability of case study findings due to the focus on a single or a few cases.

ISSN: 1696-019X / e-ISSN: 2386-3978

For the purpose of the study, two data journalism articles were selected. One utilizes data that can be characterized as big data, and the second is based on data of moderate size. Both were published by iMedLAB[3], a non-profit organization that aims to enhance transparency, credibility and independence in journalism. iMedLAB was founded in 2018 and funded by the Stavros Niarchos Foundation (SNF)[4]. The specific organization was chosen because it publishes articles in English (and in some cases, also in Greek) and datasets that have been employed in its data journalism articles. It also provides information on the process and tools employed for the collection of datasets and the creation of the visualizations, which is crucial for the purpose of this study. The two selected articles are as follows:

Article 1: MyCoast app: A data analysis of leasing agreements for beaches in Greece (published August 2nd, 2024) written by Thanasis Troboukis, available at
https://lab.imedd.org/en/analysame-ta-dedomena-gia-tis-ellinikes-paralies-tou-my-coast/

Article 2: In Greece, more people abstained from voting than participated in the European elections (published June 18th, 2024), written by Kelly Kiki and Chrysoula Marinou, available at
https://lab.imedd.org/en/more-people-abstained-from-voting-than-participated-in-the-european-elections/

3. Analysis of case studies

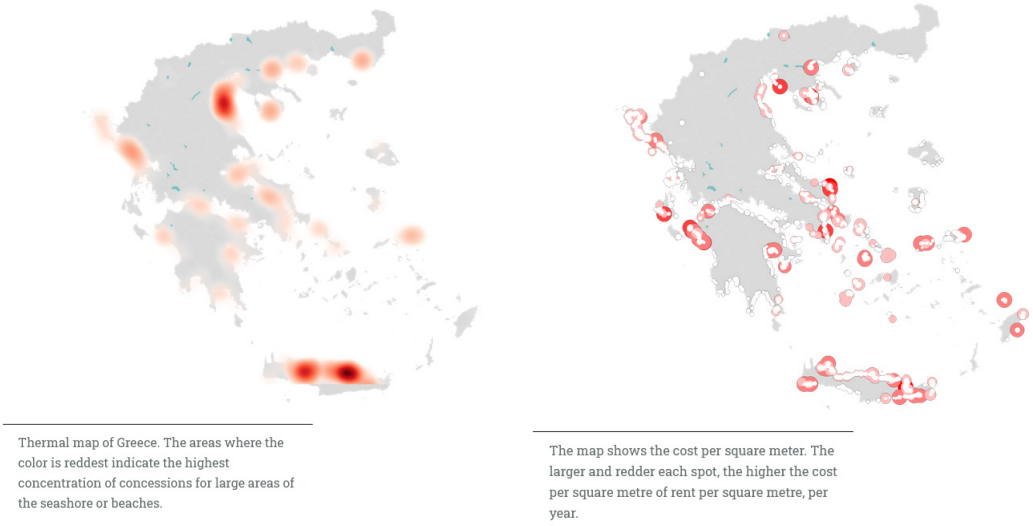
Next, the content of each article is briefly presented. Article 1 reports on a quite popular issue in Greece during the summer, using data from the MyCoast app, an application that was deployed through which Greeks could check whether the concession terms were met on the beaches they visited and, if needed, submit a report. The article includes several static and two more complex visualizations (see figure 1) that offer some degree of interactivity (the user can scroll and alter the data presented in the visualization). This article also includes a detailed map of the data used. The dataset can also be downloaded in CSV format. The inspection of the dataset revealed that it comprised 9170 rows with a total size of 4,26 Mbytes. The size of the dataset is quite large; therefore, it can be characterized as big data. The dataset was publicly available, and thus, we can assume that the dataset was obtained directly from the Greek government and no special process (i.e., web scraping, search) was utilized. Nevertheless, the iMed team studied 9,170 coastal concessions and analyzed 3,957 contracts to derive financial considerations for each lease. This means that the team managed and linked related data types. Notably, 43.0% of the records in the MyCoast app did not include relevant lease contracts. Consequently, the team implemented filtering, excluding municipalities (i.e., Paros, Mykonos, Thira, and Patmos) to ensure that the results were representative of reality. It is evident that the entire process (managing, cleaning, and filtering) was quite complex, which strongly implies the use of programming.

It is not evident whether the visualizations were created using specific tools, as they do not seem to employ template-based layouts or include any app-specific elements (watermarks, branded fonts, recognizable UI components). In contrast, visualizations have consistent and precise styling and employ simple color schemes (Figure 1). Additionally, the types of

3 https://lab.imedd.org/

4 https://www.snf.org/en/

visualizations, thermal maps, the first displaying concession concentration, and the second cost per square meter (the spot size and color relate to higher cost), are very specialized, and it is very difficult to create them with standard data analysis tools (that is, DataWrapper and Flourish). Overall, there is a high probability that the visualizations in article 1 were created via programming. It is also probable that other stages of the data journalism project (data cleaning, data understanding, and data validation) were also completed using similar methods.



Thermal map of Greece. The areas where the color is reddest indicate the highest concentration of concessions for large areas of the seashore or beaches.

The map shows the cost per square meter. The larger and redder each spot, the higher the cost per square metre of rent per square metre, per year.

Figure 1. Two static visualizations exhibiting (a) the highest concentration of concessions for large areas of the seashore or beaches and (b) the cost of leasing per square meter (https://lab.imedd.org/en/analysame-ta-dedomena-gia-tis-ellinikes-paralies-tou-my-coast/).

ISSN: 1696-019X / e-ISSN: 2386-3978

Article 2 discusses the issue of participation in the European elections in Greece. Specifically, it presents data that indicate that the abstention from the European elections in 2024 has increased in Greece by 17,3 percentage points in comparison with the previous European election in 2019, although abstention remained stable in the European Union as a whole. The article includes one static and two interactive visualizations (Figure 2) and embeds another visualization of the issue of abstention that was created by a cross-border research by EDJNet[5]. Visualizations were created using DataWrapper and Flourish. The article includes an option for readers to download the datasets used. There are four txt files and two small CSV files with the data utilized for the creation of the DataWrapper visualizations (an option that is embedded in the visualizations). As the authors of the article state, data were collected from various official online sources and scraped from PDF files (that were available online). While scraping can imply the utilization of programming, the moderate size of the data and the use of DataWrapper for the visualizations force us to assume that no programming was used for the needs of this article. Various tools that support the extraction of tables from PDFs are available and could have been employed. The same thing has probably happened with data cleaning, understanding, and validation. A spreadsheet application was probably employed, which can easily handle small data sets.

It is worth noting that although it is evident that the visualizations were created with the help of specific applications, datasets in CSV and TXT files can also be accepted in Python and R (in the case of R the process is slightly more complicated)

ISSN: 1696-019X / e-ISSN: 2386-3978

5 https://abstencao.divergente.pt/en/home

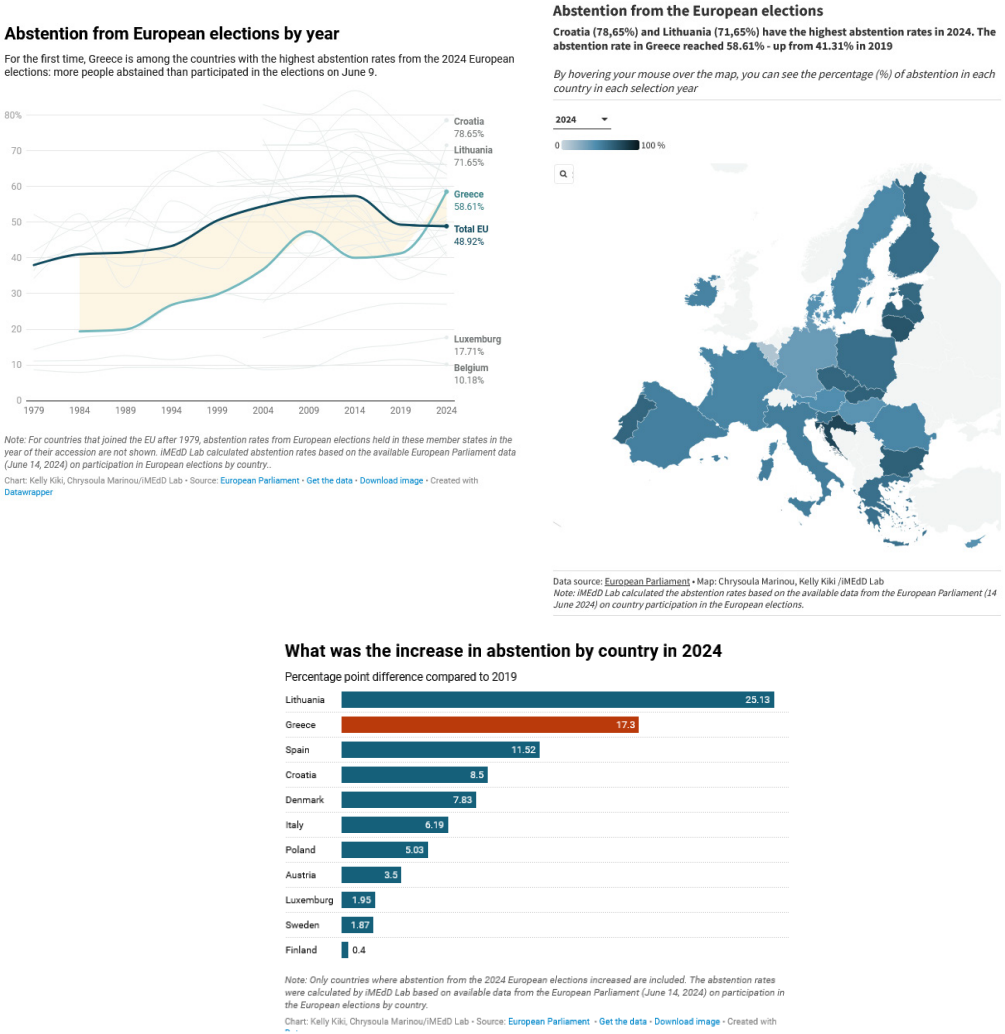


Figure 2: Visualizations (static and interactive) of various data on the issue of abstention in the European Elections (https://lab.imedd.org/en/more-people-abstained-from-voting-than-participated-in-the-european-elections/).

ISSN: 1696-019X / e-ISSN: 2386-3978

ISSN: 1696-019X / e-ISSN: 2386-3978

4. Discussion

Based on the results of the two case studies examined in the previous section, this section discusses the research questions of the study.

Concerning **RQ1,** whether coding is considered a necessity for practicing data journalism, we can answer that coding is not an absolute must, and in many cases, the tool-based approach can provide sufficient results. As is evident from Article 2 (which was presented in Section 3), the tool-based approach can be employed with good results. Heravi and Lorenz (2020) reported that although 69% of data journalists use some form of coding in their work, many successful data journalists mainly employ spreadsheet software and visual analytics tools. Borges-Rey states that the situation varies significantly between media organizations (2016). Usually, large media organizations have teams that include members with technical skills in coding and working with data, whereas in medium to small media organizations, they are more in favor of the tool-based approach. Other media organizations create teams that include both journalists with coding skills and traditional investigative reporters working with data, promoting complementary rather than mandatory coding (Parasie and Dagiral, 2013). On the other hand, Young and Hermida found an increase in no-code and low-code tools in newsrooms (2015), tools that can make complex data analysis possible for journalists without coding skills. Parasie and Dagiral view the use of coding in data journalism as an evolution in the field and not a prerequisite (2013). However, the majority of data journalists exhibit a strong interest in acquiring programming skills, as they will be able to be more productive in the field (Heravi, 2018).

In other words, while coding cannot be considered an absolute requirement for practicing data journalism, it is becoming an important skill in the field. It is worth noting that in some cases where data journalism projects involve big data, this is the only viable approach.

Moving to **RQ2**, which concerns the features of a data journalism project that influence the decision to utilize coding, we can draw some conclusions related to the creation of visualizations from the two case studies. Specifically, when specific visualizations are needed that cannot be created with the help of the available data analysis tools, the only option is to use coding. This means that if the project requires very unique visualizations, such as article 1, where specific areas need to be marked on a map, it is quite difficult or even impossible to find data analysis tools that can create such a visualization. However, in article 2, the visualizations are not very specialized (line chart, bar chart, and choropleth map); thus, well-known data analysis tools have been employed. Nevertheless, such visualizations can also be created by coding, as both programming languages (Python and R) offer extensive libraries and packages for creating such visualization types.

As for the other stages of development of the data journalism projects in our case studies, there is no indication of what approach has been employed. Nevertheless, we can make some speculations. Specifically, in Article 1, the data reside in a single PDF file, which, although quite extensive in size, is still manageable by regular spreadsheet applications (Microsoft Excel and Google Spreadsheet). However, because the visualizations were probably created with programming, there is a high probability that other stages of development (for example, data cleaning, data understanding, and data validation) were also completed through coding. In the case of article 2, the data employed were available from four different files (with relatively small sizes). Because the datasets were obtained from different sources and, as mentioned, scraped from PDF files, there is some probability that a programming approach was used. However, if we consider the facts that a) there are several tools that

can support this process, and b) the visualizations were created by data analysis apps, it is more probable that no programming was employed for the needs of article 2. The size of the dataset is a very important factor in selecting a suitable approach (tool-based or code-based). Additionally, visualization requirements are an important factor. It is also worth noting that it is not very common to use a mixed approach (tool-based and code-based) during a project.

In addition to the data journalism features that were examined in our case study analysis, there are other parameters that affect the selection of the approach in developing a data journalism project, that derive from the scientific literature. Personal interest in technology, prior exposure to programming concepts, and perceived career benefits are important parameters influencing coding adoption by journalists (Appelgren & Nygren, 2014). Organizational context (newsroom size, available resources, and institutional support) is significantly responsible for selecting the code-based approach (Fink & Anderson, 2015). Similarly, Joung and Hermida highlighted the role of existing technical infrastructure and workflow integration as positive parameters for embracing the code-based approach. Coddington emphasized the factor of working with large datasets and the requirement of frequent data updates as a reason for choosing to drop the programming approach (2015).

5. Conclusions

This study examined the utilization of programming in modern data journalism. The study's conclusions reveal the relationship between coding and practicing data journalism. First, while coding is not a requirement for data journalism, it has become a valuable skill in the field. The analysis shows that both tool- and code-based approaches can work in many cases for data journalism projects. The choice between the two depends on many factors, including project complexity, dataset size, repetition needs, and visualization requirements.

Specifically, the case study analysis shows that the decision to code is mainly driven by the size and complexity of the data being analyzed, with larger datasets benefiting more from a code-based approach. This decision is also driven by the specificity of visualization requirements, where unique or complex visualizations require coding solutions. Other factors include workflow integration and automation, available resources and technical infrastructure, and journalists' technical background and skills. Most importantly, there is no one-size-fits-all approach to data journalism. Coding can provide more flexibility and capabilities for complex data analysis and visualization, but traditional tool-based approaches are still viable and work for many data journalism projects, especially those with moderately sized datasets and standard visualization needs. New tools and platforms, such as Google Colab and Excel's Python integration, are gradually breaking down the technical barriers to coding in journalism. This means that the distinction between coded and non-coded approaches will become less clear as hybrid approaches become more accessible to journalists with less technical backgrounds. These findings have significant implications for journalism education and newsroom practices. This means that while coding skills are useful, they should be seen as a complementary, not mandatory, tool in the data journalist's toolkit. The focus should be on strong journalistic principles and storytelling, with technical skills serving to enhance, not replace, the core of the profession.

However, it is worth noting that this study exhibits certain limitations, since it includes a small sample of case studies from one media organization. Additionally, it is not easy to define with complete certainty the approach used for each data journalism article without acquiring the necessary information from the actual journalism team that developed the projects. Nevertheless,

ISSN: 1696-019X / e-ISSN: 2386-3978

the current paper should be treated as a concept study, with the aim of opening a debate about the role of programming in modern data journalism.

Future research could focus on the long-term impact of programmatic literacy in newsrooms, the effectiveness of hybrid approaches combining traditional and coded methods, and the evolution of data journalism as new tools and technologies emerge by conducting surveys of data journalists. This could provide the scientific community with insights into how data journalism continues to adapt and evolve with technological advancements and changing newsroom needs.

ISSN: 1696-019X / e-ISSN: 2386-3978

6. Acknowledgements

This article has been translated into Brian O'halloran, to whom we are grateful for his work.

7. Specific Contributions of Each Author

	Name and Surname(s)
Conception and Study Design	Andreas Veglis
Methodology	Andreas Veglis
Data Collection and Analysis	Andreas Veglis
Discussion and Conclusions	Andreas Veglis
Writing, Formatting, and Revision	Andreas Veglis

8. Conflicts of Interest

The author declares that they have no conflicts of interest.

9. Bibliographical references

Angelou, I., Kansaras, V., Kourkouridis, D., & Veglis, A. (2020). Social media followership as a predictor of news website traffic. *Journalism Practice*, *14*(6), 730-748.

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden: Introducing new methods and genres of journalism into 'old' organizations. *Digital Journalism*, 2(3), 394–405.

Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2020). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, 21(7), 950-973.

Balaji, N., Pai, B. K., Bhat, B., & Praveen, B. (2021, February). Data visualization in splunk and Tableau: a case study demonstration. In *Journal of Physics: Conference Series* (Vol. 1767, No. 1, p. 012008). IOP Publishing.

Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A data visualization tool. *The Journal of Economic Education*, *51*(3-4), 317-328.

Bauzon, J., Murphy, C., & Wahi-Gururaj, S. (2021). Using macros in Microsoft Excel to facilitate cleaning of research data. *Journal of Community Hospital Internal Medicine Perspectives*, 11(5), 653-657.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544-559.

Borges-Rey, E. (2016). Unravelling data journalism: A study of data journalism practice in British newsrooms. *Digital Journalism*, 4(3), 336-352.

Bounegru, L., Gray, J., Venturini, T., & Mauri, M. (2018). A field guide to'Fake News' and other information disorders. *A Field Guide to" Fake News" and Other Information Disorders: A Collection of Recipes for Those Who Love to Cook with Digital Methods*, Public Data Lab, Amsterdam (2018).

Bradshaw, P. (2018). *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam University Press.

Burke, J. (2023). Why and how to use Google Colab. T*echTarget* Available at https://www.techtarget.com/searchenterpriseai/tutorial/Why-and-how-to-use-Google-Colab

Cairo, A. (2012). The Functional Art: An introduction to information graphics and visualization. New Riders.

Card, S. K., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.

Coddington, M. (2015). Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331–348.

Colman, F., Bühlmann, V., O'Donnell, A., & van der Tuin, I. (2018). *Ethics of coding: A report on the algorithmic condition*. Available at https://ualresearchonline.arts.ac.uk/ id/eprint/14236/1/Ethics_of_Coding_A_Report_on_the_Algorit.pdf

Crawley, M. J. (2012). *The R book*. John Wiley & Sons.

Cruz, J. (n.d.). *Why journalists need data and coding skills more than ever*. International Journalists' Network. Available at https://ijnet.org/en/story/why-journalists-need-data-and-coding-skills-more-ever

Few, S. (2013). 35. Data visualization for human perception. *The Encyclopedia of Human-Computer Interaction,* International Design Foundation.

Fink, K., & Anderson, C. W. (2015). Data Journalism in the United States: Beyond the "usual suspects". *Journalism Studies*, 16(4), 467-481.

Gray, J., Bounegru, L., & Chambers, L. (2012). *The Data Journalism Handbook*. O'Reilly Media.

Guerrero, H., Guerrero, R., & Rauscher. (2019). *Excel data analysis*. Springer International Publishing.

Hamilton, J. T. (2016). *Democracy's Detectives: The Economics of Investigative Journalism*. Harvard University Press.

ISSN: 1696-019X / e-ISSN: 2386-3978

Hannaford, L. (2015). Computational journalism in the UK newsroom. *Journalism education*, 4(1), 7-21.

Heravi, B. R. (2018). Data journalism in 2017: A summary of results from the global data journalism survey. In T*ransforming Digital Worlds: 13th International Conference,* iConference 2018, Sheffield, UK, March 25-28, 2018, Proceedings 13 (pp. 107-113). Springer International Publishing.

Heravi, B. R., & Lorenz, M. (2020). Data journalism practices globally: Skills, education, opportunities, and values. *Journalism and media*, 1(1), 26-40.

Kinnestand, S. (2023). Introducing Python in Excel: The Best of Both Worlds for Data Analysis and Visualization, Microsoft 365 Blog. Available at https://rb.gy/nshr7e.

Knight, M. (2015). Data Journalism in the UK: a preliminary analysis of form and content. *Journal of Media Practice*, 16:1, pp. 55-72.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34(1), 26-38.

Lewis, S. C., & Westlund, O. (2015). Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3(3), 447–466.

Lorenz, M. (2010). Data driven journalism: What is there to learn? Presented at *IJ-7 Innovation Journalism Conference*, 7–9 June, Stanford, CA.

Loth, A. (2019). *Visual analytics with Tableau*. John Wiley & Sons.

Lutz, M. (2001). *Programming python*. O'Reilly Media, Inc.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society*, 15(6), 853-871.

Porlezza, C. (2023). The datafication of digital journalism: A history of everlasting challenges between ethical issues and regulation. *Journalism*, 25(5), 1167-1185.

Protopsaltis, A., Sarigiannidis, P., Margounakis, D., & Lytos, A. (2020, August). Data visualization in internet of things: tools, methodologies, and challenges. In *Proceedings of the 15th international conference on availability, reliability and security* (pp. 1-11).

Rogers, S. (2008). Turning official figures into understandable graphics, at the press of a button. Inside the Guardian Blog, https://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata

Ruppert, E., Isin, E., & Bigo, D. (2017). Data politics. *Big data & society*, 4(2), DOI: 10.1177/2053951717717749.

Setiyanto, S., & Setiawan, I. (2022). Data science with excel. *International Journal of Computer and Information System* (IJCIS), 3(3), 104-110.

Siapera, E., & Veglis, A. (Eds.). (2012). *The handbook of global online journalism*. John Wiley & Sons.

ISSN: 1696-019X / e-ISSN: 2386-3978

Silverman, C. (2014). Verification Handbook. European Journalism Centre http://verificationhandbook.com/downloads/verification.handbook.pdf.

Simon, B. (2021). Coding With Data in the Newsroom, in Bounegru, L., & Gray, J. (eds). *The Data Journalism Handbook: Towards a Critical Data Practice* (p. 124-127). Amsterdam University Press.

Tellis, W. (1997). Application of a case study methodology. *The qualitative report*, 3(3), 1-19.

Tock, K. (2019). Google CoLaboratory as a platform for Python coding with students. *RTSRE Proceedings*, *2*(1).

Veglis, A. (2013). Education of Journalists on ICTs: Issues and Opportunities. *Journal of Applied Journalism & Media Studies*, 2(2), 265-279.

Veglis, A., & Bratsas, C. P. (2021). Data Journalism: Definition, Skills, Difficulties, and Perspectives. In *Encyclopedia of Information Science and Technology*, Fifth Edition (pp. 1140-1151). IGI Global.

Veglis, A., & Bratsas, C. (2017a). Journalists in the age of Data Journalism: The case of Greece. *Journal of Applied Journalism & Media Studies*, *6*(2), 225-244.

Veglis, A., & Bratsas, C. (2017b). Towards a taxonomy of data journalism. *Journal of Media Critiques*, *3*(11), 109-121.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage publications.

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism? A content analysis of data journalism awards finalists 2012–2015. *Journalism practice*, 12(1), 115-135.

Young, M. L., & Hermida, A. (2015). From Mr. and Mrs. Outlier to Central Tendencies: Computational journalism and crime reporting at the Los Angeles Times. *Digital Journalism*, 3(3), 381-397.

Wu, S. (2013). A review on coarse warranty data and analysis. *Reliability Engineering and System*, 114, 1-11.

ISSN: 1696-019X / e-ISSN: 2386-3978