



**European  
Journalism  
Centre**

# **The State of Data Journalism Survey 2023**

The most comprehensive annual data journalism survey.



**DataJournalism.com**



# Table of Contents

Our Mission	2
Methodology	3
Key Takeaways	5
Demographics	7
Gender	7
Age	7
Country of work	8
Education	9
Employment	11
Occupation	11
Years of experience	12
Learning data journalism	13
Income	14
Company	16
Skills and Tools	17
Skill level	17
Programming	18
Graphic tools	20
Tools	20
Work Practices	21
Type of data journalism	22
Work tasks	24
Scope	25
Beat	26
Medium	27
Data used	28
Dedicated data unit	29
Projects	30
Collaborations	30
Challenges	32
Challenges to producing more	32
Data access and quality	33
Barriers to learning	34
Value of data journalism	34
AI and OSINT	36
AI	37
Conclusion	42
Thank you	43



# Our Mission

This report presents the findings of The State of Data Journalism 2023 Survey, highlighting key trends and changes in demographics, education, employment, skills, and challenges faced by professionals in the field. Before delving into each section in-depth, below is a mission statement about our project.

Data journalism is a powerful force in the ever-evolving media landscape, pushing the boundaries of storytelling and captivating audiences when attention is the currency defining success. But it's not just about flashy visuals— data journalism tackles some of today's hardest and most important journalistic tasks: fact-checking and evidence-gathering.

The State of Data Journalism Survey exists to **keep track of the latest developments in this exciting and fast-moving field** and shed light on where the profession is heading - and where it faces hurdles - to evolve and flourish.

In 2021 and 2022, we rolled out in-depth data journalism surveys. What did our findings show? **Significant and consistent gender imbalances and data quality issues** vary depending on location. Those elite dedicated data teams? Scarce, especially outside the legacy media giants. Oh, and the pandemic? It played a role in bringing a third of our respondents into the data journalism world.

At its core, data journalism is about **blending data, tools, and visuals to craft impactful stories**. But here's the real challenge—data journalists must have strong reporting skills combined with the ability to work with stats, design visuals and even wield a bit of code in certain programming languages. Our surveys revealed a thirst for levelling up these skills.

With the 2023 survey, we seek to continue our tradition of understanding and explaining the data journalism profession to our global audience. This year we are introducing two new sets of survey questions about Open-Source Intelligence and Artificial Intelligence.

This study reflects Datajournalism.com and EJC's commitment to continuously support the data journalism community. Join us in our mission to monitor and map the state of data journalism today.



# Methodology

The State of Data Journalism 2023 Survey comprised a total of 76 questions and had 776 respondents, of which 768 were used for the analysis. The survey was organised into seven sections: demographics, industry structure, skills and tools, work practices, challenges and opportunities, and the usage of OSINT and AI. It was available in three languages: English, Italian, and Spanish.

## 1. Population of interest and sampling strategy

The population of interest was the global community of individuals involved in data journalism. Targeted respondents included full-time and part-time employed data journalists, freelancers, data editors and team leads, trainers, faculty members, educators, and students. In the absence of a directory of people who are part of the data journalism industry in 2023, we discarded the possibility of drawing a random or representative sample. Instead, we followed the approach of trying to reach respondents as widely as possible through a variety of channels.

## 2. Outreach strategy and incentives

The survey was open between December 7th, 2023 and January 16th, 2024. Participation was encouraged through various communications channels to minimise bias obtained by targeting solely one online community. We used direct mailing, social media promotion and asked the DataJournalism.com and European Journalism Centre communities to help spread the word.

## 3. Survey logic

Questions targeting a specific subgroup were only shown to those respondents to minimise survey length while maximising survey inclusion. Still, questions about journalistic practices were left open to all (this was done to reflect that students, educators, and editors might be involved in producing and publishing journalistic data work from time to time). Most survey questions were optional.



#### **4. Data Cleaning**

For the analysis, only complete questionnaires were retained. We included all questionnaires where respondents had filled in all of our mandatory questions. However, we obtained near to full responses throughout the entire questionnaire (the median response rate was 99%). We found eight duplicate names and email addresses. We randomly selected which questionnaire to keep for each of these duplicates.

#### **5. Metadata**

The reported statistics have a 95% confidence level and a margin of error of  $\pm 4$  percentage points.

With the survey in its third edition, we asked participants whether they had previously participated in the survey: 13% of respondents took the survey in 2021, and 22% took the survey in 2022. Of these, 7% have taken both previous editions of the survey.

85% of respondents took the survey in English, 8% in Spanish, and 7% in Italian. 58% of respondents found the survey length good, while 32% found it a bit long, which is an overall improvement from the 2021 and 2022 editions (41% and 37%, respectively). We appreciate the feedback, and we will continue to improve our questionnaire.



# Key Takeaways

**Demographics** The survey reveals a balanced gender distribution in data journalism, with nearly equal representation of men and women, signifying a **significant shift towards gender inclusivity** in the industry. The prevalence of younger professionals, especially in the 25-34 age group, points to a trend of **emerging talent** shaping the future of data journalism.

---

**Occupation** A majority of data journalists are employed full-time in news companies or organisations, indicating a **stable job market** within traditional media settings. The increase in full-time freelancers from 2021 to 2023 suggests a **growing trend towards flexible work arrangements** in the industry.

---

**Experience and skills** Data journalists predominantly possess strong journalism skills. Still, there's a **notable gap in technical skills** like data analysis and visualisation, underscoring the need for comprehensive skill development in these areas. The correlation between experience in data journalism and higher skill levels emphasises the **value of ongoing professional development** and experience in the field.

---

**Data** Public official governmental data is the most commonly used type of data, highlighting its crucial role in data journalism. The **use of FOI-obtained data** being **subject to regional variations** demonstrates differences in data sourcing practices and the importance of legal frameworks in accessing information.

---

**Projects** Most data journalism projects are medium-term endeavours, completed within weeks or months, indicating the **time-intensive nature of this work**. The predominance of small team collaborations in these projects reflects the **collaborative and interdisciplinary nature of data journalism**.

---



---

**Dedicated data unit** The presence of dedicated data units in a quarter of the respondents' organisations, especially in medium to large ones, underscores the **growing recognition of specialised data journalism teams** within the broader media landscape. The trend towards smaller data units suggests a **focus on agility and specialisation** within these teams.

---

**Challenges** Access to quality data, time pressure, and lack of data analysis knowledge are the top challenges faced by data journalists, highlighting the **need for better data accessibility, efficient workflow management, and skill enhancement** in the industry.

---

**AI and OSINT** A small portion of data journalists have incorporated AI and OSINT into their work, primarily for **content search and verification**, signalling a slow integration of advanced technologies in journalistic practices to deal with content-related challenges. The challenges faced in using AI, such as limited understanding and concerns about bias, point to the **need for greater education and ethical guidelines in the application of these technologies** in data journalism.

---



# Demographics

## Gender

In 2023, 49% of respondents identified as men, 48% as women, 1% as non-binary / genderqueer, and 1% preferred not to answer. The near parity between male and female respondents marks a significant shift from 2022, reflecting a progressive gender balance in the industry.

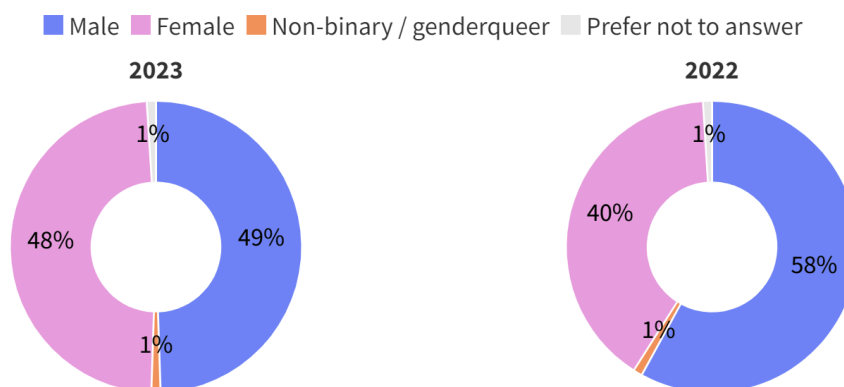


Figure 1: The gender distribution in the 2023 and 2022 surveys.

## Age

Over 69% of respondents are younger than 44 years of age, compared to 60% in 2022. The largest group (33%) is aged between 25-34. Compared to previous years, the distribution has shifted in favour of younger cohorts. The younger age trend, with a notable increase from 2022, suggests a dynamic industry attracting emerging talent.



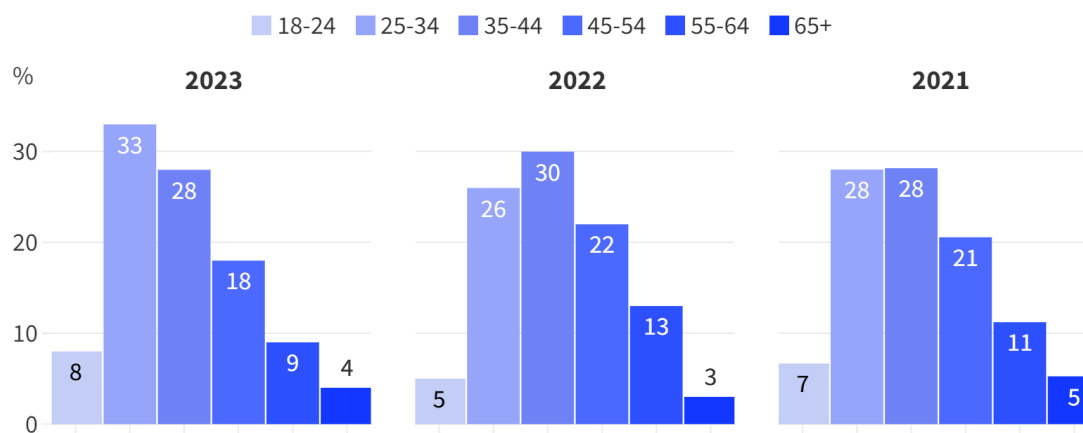


Figure 2: The age distribution in the three survey editions.

Across gender, the largest age group is the same (25-34). However, it is much higher for non-binary/genderqueer respondents and female respondents than for men. This implies that there is a diverse gender representation in younger cohorts, while older age groups have an overrepresentation of men.

## Country of work

The United States has the highest share of people in data journalism (8%). This is followed by Italy (7%) Germany (6%), the United Kingdom (5%), and Russia (5%). Compared to previous years, the US dominance is lessening, indicating a geographical diversification in data journalism. Regionally, Brazil has the highest response rate for a Latin American country (1%). In Africa, the countries with the highest number of respondents include Nigeria (4%) and Kenya (3%).

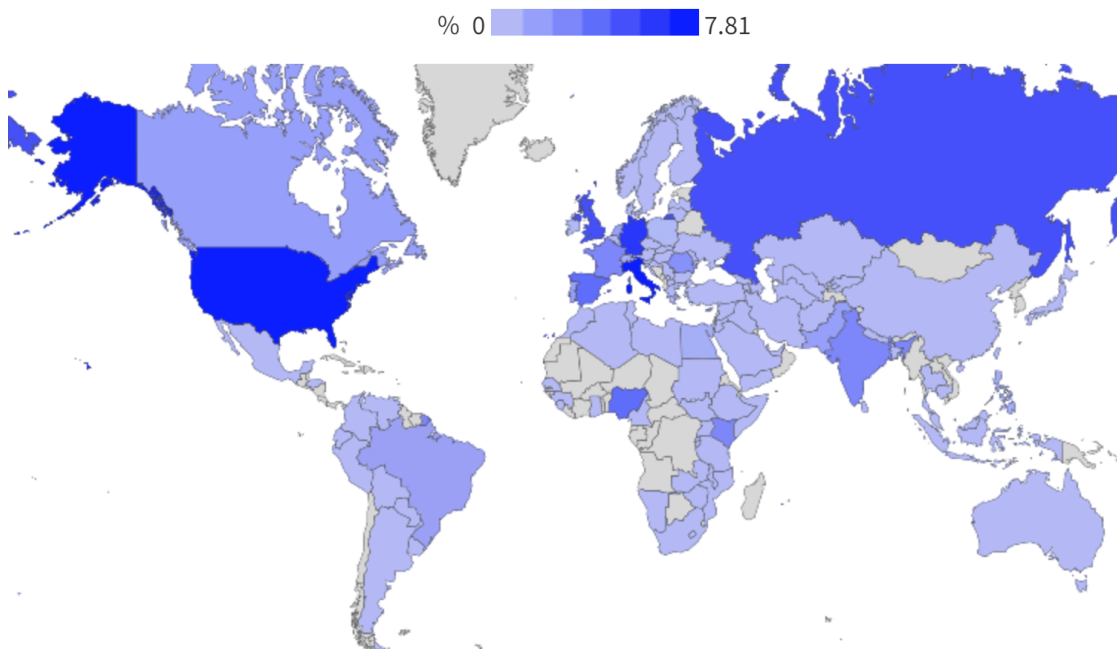


Figure 3: The country distribution in the 2023 survey.

Amongst countries with a minimum of 10 respondents, Pakistan has the highest share of men at 94%. On the opposite spectrum, Poland's respondents are 89% female. Canada tops the ranking for non-binary/genderqueer people (14% of country respondents), followed by Czechia (12%). These trends highlight diverse gender dynamics in data journalism across different regions.

## Education

Master's degrees are the most common type of education completed by people in data journalism (46%). Like in previous years, women tend to be the most highly educated, with 59% having obtained a Master's degree or PhD (against 54% in the overall data set). The high educational level, especially among women, suggests an increasingly skilled workforce, potentially driving the industry towards more sophisticated and diverse journalistic practices.

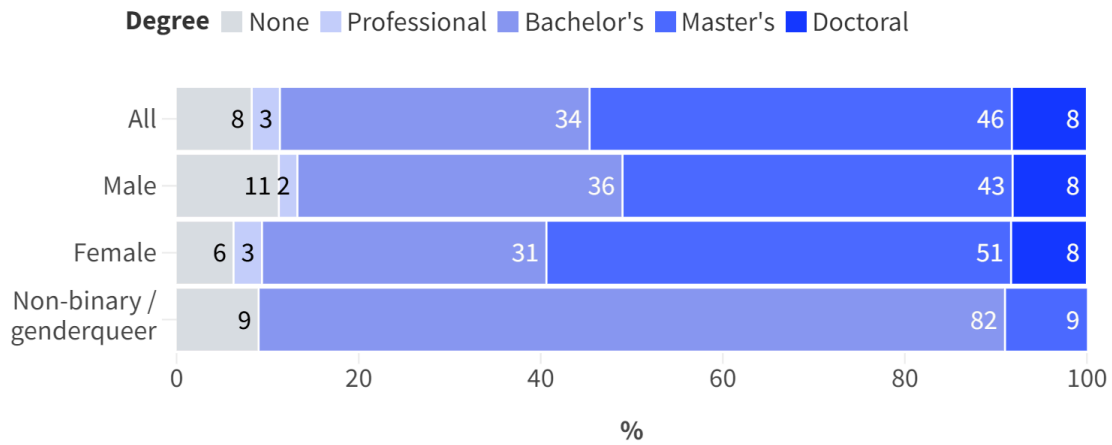


Figure 4: The education level distribution in the 2023 survey by gender.

Regardless of the type of degree, Arts and Humanities are by far the largest disciplines in which data journalists specialise. Doctoral degrees have a higher share of Social Sciences (26%) graduates, while more Master's are pursued in Business, Finance, or Law (8%). On the other hand, Professional Degrees are very rarely pursued in the Social Sciences (9%) and otherwise show a very high proportion of Formal Sciences or "Other" (13% respectively).



# Employment

## Occupation

As in 2021 and 2022, the primary occupation in data journalism in 2023 is full-time employment at news companies or organisations (36%). Another trend that is consistent with the 2022 survey results is the wide gap between full-time and part-time in employment contracts (a difference of 30 percentage points), where the same difference is not found among freelancers (a difference of 3 percentage-points, with 13% working full-time and 10% working part-time). Students make up 7% of respondents. The 2023 sample includes fewer educators (9%) compared to 2022 (14%) and 2021 (15%). An increase in full-time freelancers from 10% in 2021 to 13% in 2023 might suggest a shift towards more flexible work arrangements.

As we break down occupation by gender, we find the previously noted gap in Editor/Team Leading positions between men and women (4 percentage-points in 2022) disappearing, with both cohorts at 8%. The closing gender gap in leadership roles suggests progress towards gender equity in higher positions.

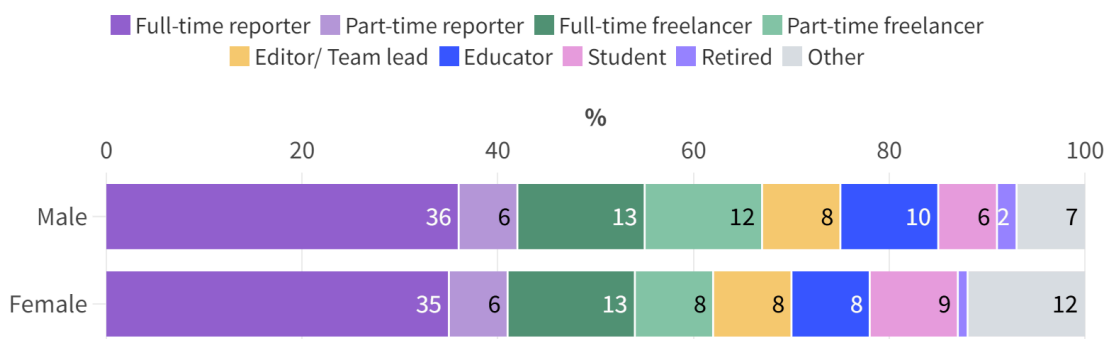


Figure 5: The occupation breakdown in the 2023 survey, by gender.

Students are by far the youngest in the industry as 77% are under 34 years of age. On the opposite end of the spectrum, relatively few educators are under 45 years



of age (44%). Full-time freelancers tend to be older than full-time employees (37% of full-time freelancers are over 35 years of age, against 21% of full-time employees). In contrast, editors are normally distributed around the largest group being 35-44 years of age (47%).

In terms of within-country trends, France has a higher proportion of full-time freelancers (35%) compared to the share across the whole dataset, while Italy of students (16%).

Looking at the geographical distribution of different occupations, 17% of employed data journalists work in the United States. The United States also has the most educators (12%).

## Years of experience

The largest group (30%) of respondents has between 3-5 years of experience in data journalism. Similarly to 2021 and 2022, in 2023 there are more respondents who have been in the industry for less than three years (35%) than for more than five years (30%).

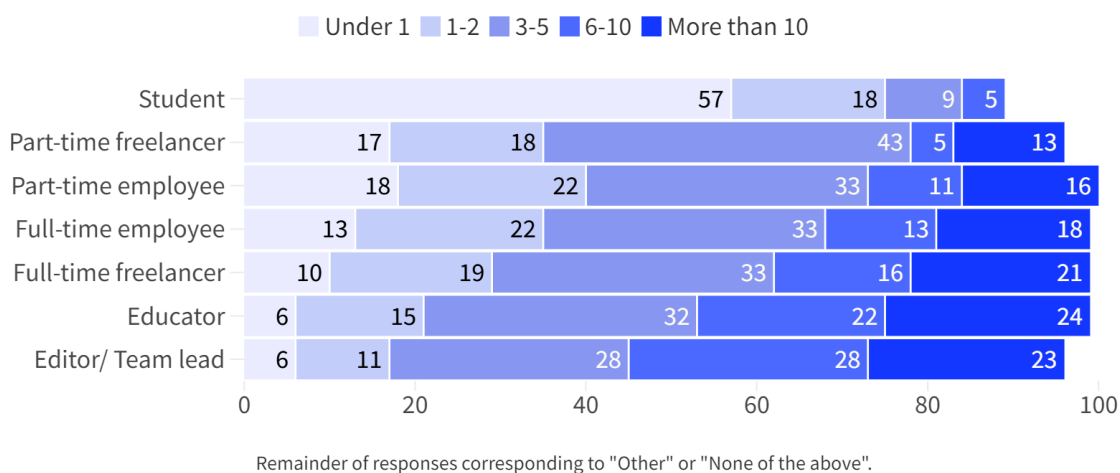


Figure 6: The number of years in the industry from the 2023 survey, by occupation.

Editors have the longest experience in data journalism, with 51% having been in the industry for more than six years, followed by educators (46%). Part-timers,



regardless of whether they are freelancers or contracted reporters, tend to have less experience than their full-time counterparts. Freelancers tend to have been in the field longer than contracted reporters.

By gender, women tend to have been in the field for less time, with 42% having less than three years of experience, against 28% of men. Overall, the experience distribution suggests a maturing industry but also highlights the age and experience disparity between genders in this industry.

## Learning data journalism

When it comes to learning the profession, the most common means is self-taught learning through online resources (53%). Only just under a quarter of respondents learned data journalism through higher education. Overall, the percentages are pretty stable over time, with the exception of self-taught, which, both in the online and offline typology, sees a decrease of around 10 percentage points between 2022 and 2023. Independent learning is highly common across all professions. On the other hand, workplace training is more common among editors and employees, but even among these groups, only a third has received workplace training in data journalism. Comparatively, these professions are likelier to have learned data journalism through formal educational paths, such as online certifications and training events outside the organisation (around half of the individuals in these groups). Formal education is, in fact, the second most common formation type after self-learning for contracted individuals as well as freelancers. For instance, around half of freelancers said they attended bootcamps, training events, or obtained professional certifications.

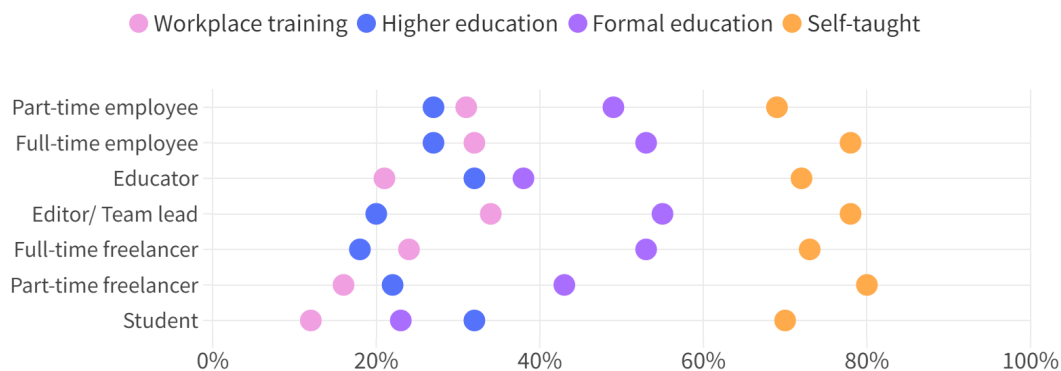


Figure 7: How respondents learn data journalism, by occupation.



In 2023, 23% of respondents are solely self-taught in data journalism, compared to 35% in 2022. The decrease in purely self-taught professionals could reflect the growing need or availability for structured learning. Despite the drop, this means that nearly a quarter of people in the industry are self-taught in data journalism. These individuals are more likely to be freelancers or students, where around one in three are solely self-taught. Self-learners tend to have fewer years of experience in data journalism than those who have undertaken workplace training or higher education in data journalism.

## Income

95% of respondents in 2023 provided information about their income. Of those, over half earn less than \$25K annually, regardless of occupation. The share of respondents in this group has steadily increased over time, starting off at 44% in 2021, up to 47% in 2022, and 52% in 2023. Similarly, while 12% of respondents to this question earned more than \$75K in 2021 and 2022, in 2023, the equivalent figure is 8%, potentially reflecting an influx of entry-level professionals or saturation of lower-paying roles.

By gender, women's salaries are often clustered in the mid-low range, while men's salaries are more spread out, meaning there are more men that earn \$1 and \$9K compared to women, but there are also more men that earn over \$75K. These patterns are consistent with the results from the 2022 survey and hint at gender disparities in earning potential within the industry.

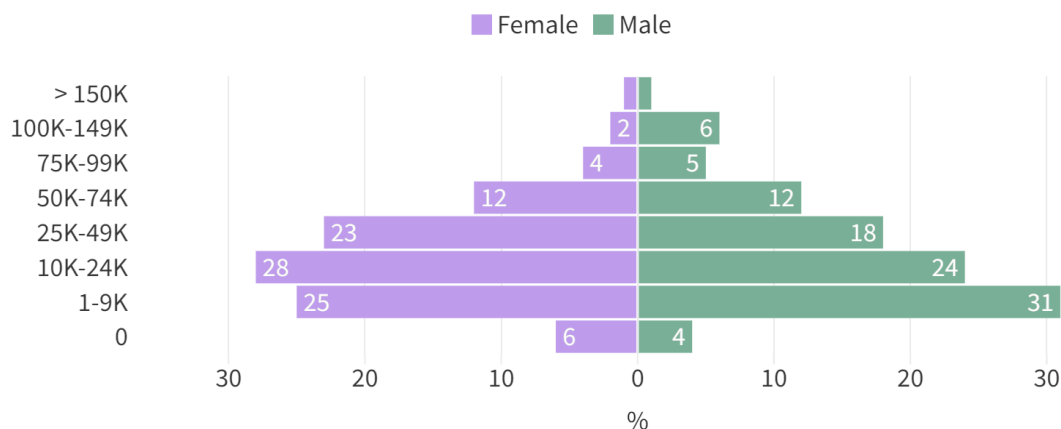


Figure 8: The income distribution in the 2023 survey, by gender.



Full-time employment, education roles, and leading roles offer the highest compensation in the industry in 2023 as they did in 2022. By company size, the highest salaries are found within the largest organisations. Nearly one in four from organisations with 500+ employees earns over \$50K. Income distributions by country vary significantly. Workers in the United States earn more, followed by the UK and Germany. However, these figures are not adjusted by cost of living.

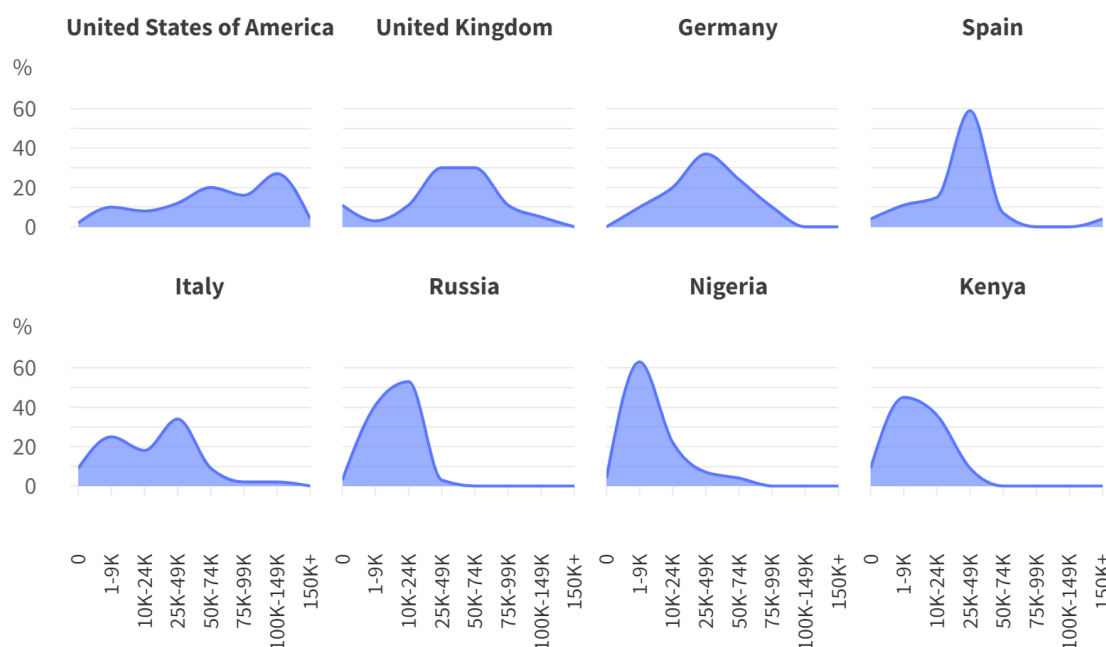


Figure 9: The income distribution in the 2023 survey, by country.

Top earners are found in a few countries. Alongside the US, those who earn over \$100K work in Qatar, Switzerland, Denmark, Singapore, Saudi Arabia, Canada, Austria, Australia, South Africa, Spain, Slovenia, and New Zealand.



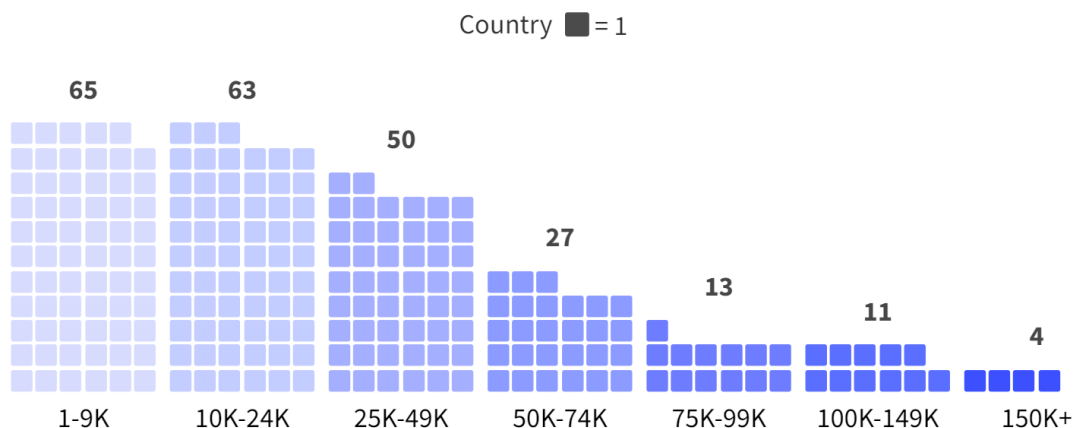


Figure 10: The number of countries by income bracket.

## Company

Nearly half of respondents work for one organisation. However, compared to 2022, the number of individuals who work for two organisations or more is on the rise (35% versus 28%), suggesting a dynamic, perhaps fragmented, job market in data journalism. Freelancers predominantly work for more than two organisations (56% of full-time freelancers and 38% of part-time freelancers). Only 12% of full-time freelancers work for only one organisation. On the other hand, 72% of full-time employees and 49% of part-time employees work for one organisation.

There is no clear distribution when it comes to the size of the company respondents operate in. Overall, just over one in five works in companies of 500+ individuals, and the remainder of respondents are quite evenly spread out. These figures are stable with results from previous survey editions and point to varied opportunities in different organisational contexts.

The United States, the United Kingdom, and Germany are the three countries where the largest share of the workforce operates at companies of 500+ employees, and the second largest share is of individuals working alone. In comparison, small to medium sized enterprises are more common in Italy, Spain, and France.



# Skills and Tools

## Skill level

Journalism is the area in which data journalists feel they most excel, with machine learning on the opposite end of the spectrum. While 56% consider themselves advanced at journalism, only 12% say they are advanced in data analysis and 15% in data visualisation. As a result, the previously observed gap between data skills and journalistic skills persists in 2023.

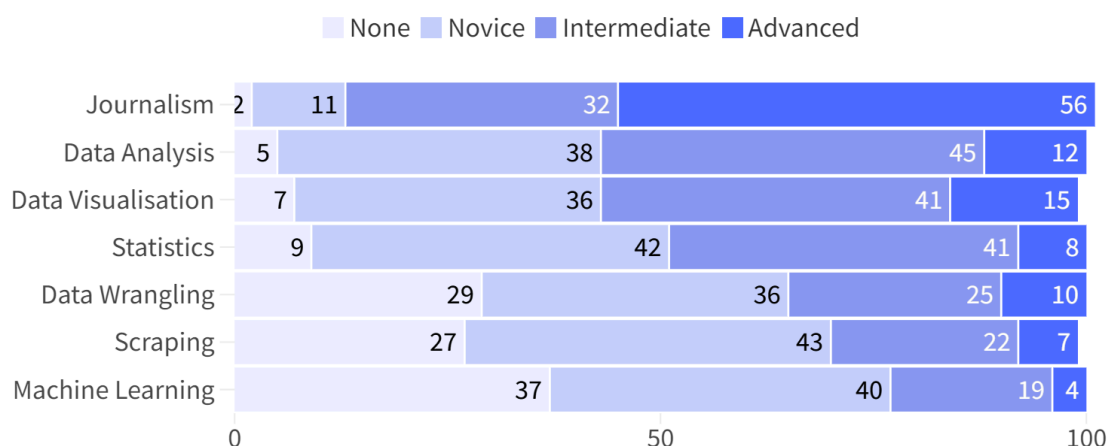


Figure 11: Skill level by occupation.

When breaking it down by occupation, most profiles follow the same trends in terms of rating skill level. Overall, students tend to be the most inexperienced in nearly all areas. As in 2022, the biggest gap in proficiency between students and industry is in journalism, where only 16% of students have rated their skill level in journalism as advanced.

Like in 2022, experience in data journalism positively correlates with higher skill level rating, and this is particularly the case in data analysis and data visualisation. The correlation between experience in data journalism and higher skill levels emphasises the value of ongoing professional development and



experience in the field. The areas with least improvement over time are machine learning and statistics. These are also amongst the skills where people in the industry would most like to advance. Similar to previous survey editions, journalism is the area in which most respondents have received training (56%). The largest gap where respondents have received training and desire upskilling is in machine learning, where only 14% have been trained on this skill, while 57% desire upskilling in this area. The second and third largest gaps are in data wrangling and web scraping. The interest in upskilling, particularly in areas like machine learning, highlights a growing awareness of the need to adapt to emerging technological trends.

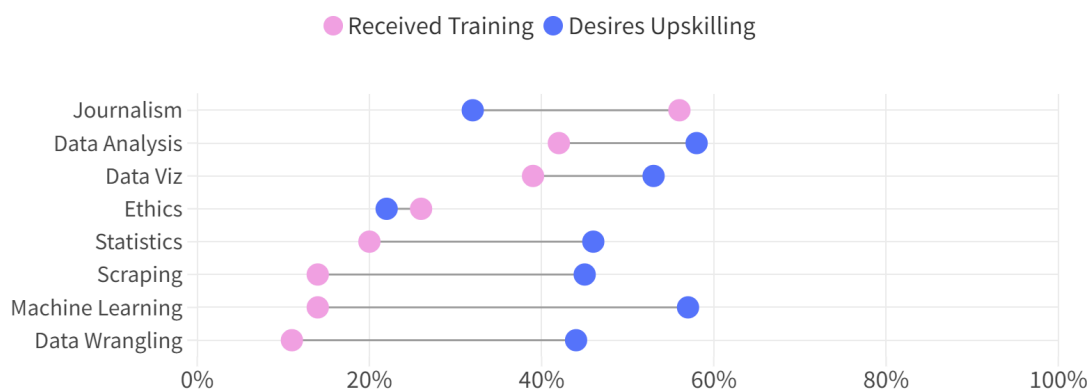


Figure 12: Having received training and desiring upskilling, by skill.

## Programming

Among those who code as part of their work (19% in 2023), the share of those who use Python is increasing (69%), as is that which uses HTML and CSS (61%) or Javascript (46%). On the other hand, fewer respondents use R in 2023 compared to 2021 (from 46% to 39%). The increasing usage of Python, HTML/CSS, and JavaScript among those who code in their work suggests a trend towards languages and technologies that are versatile and have strong community support. Python's rise could be due to its simplicity and the extensive libraries available for data analysis, machine learning, and visualisation. HTML/CSS's growth can be attributed to the need for data journalists to present their findings on the web attractively and interactively. The slight decrease in R



usage (too small to be deemed significant) might be because Python is increasingly able to perform many of the statistical analysis tasks traditionally done in R.

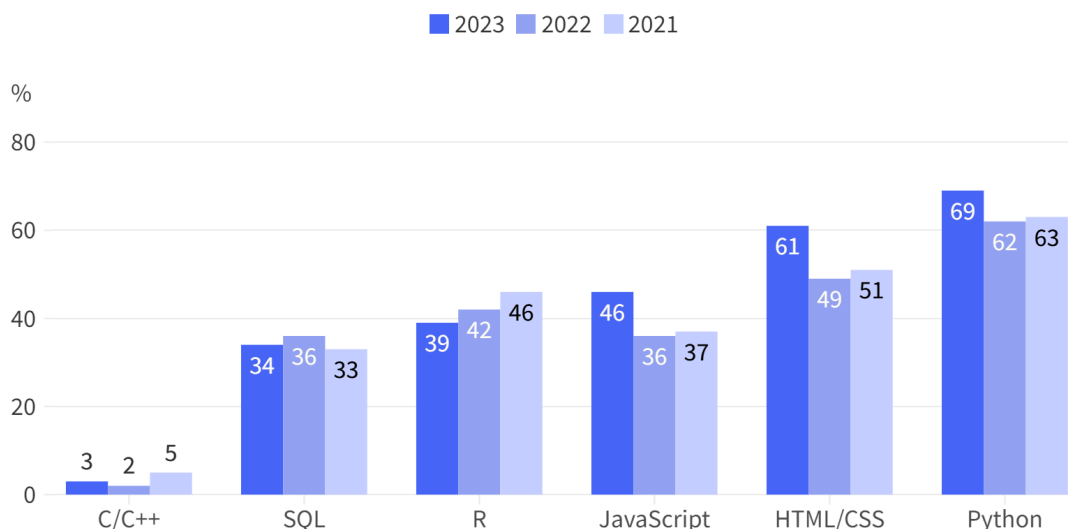
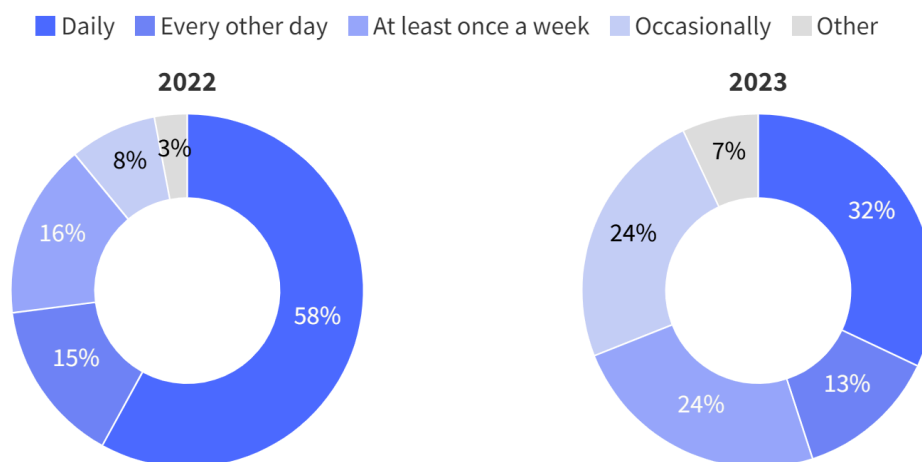


Figure 13: The usage rate of programming languages among programmers, over time.

Coding is a skill mostly acquired through self-learning. However, over one in three learns how to programme while pursuing higher education studies. Rarely respondents have learned how to code at work (16%). In terms of experience, in 2023 64% of programmers have known how to code for more than six years, against 46% of respondents in 2022.

The share of those who use coding on a daily basis is increasing rapidly (58% in 2023 against 32% in 2022). While around half of respondents used to programme once a week or occasionally in 2022, this share is down to less than a quarter of programmers in 2023. These shifts might indicate that coding is becoming an integral part of the daily work for some data journalists, possibly due to the growing demand for timely data analysis and the need for rapid development of data-driven stories. This shift may also reflect the increasing ease with which journalists can use programming within some aspects of their work.



This question was shown to those who answered they code or develop applications as part of their job.

Figure 14: The frequency of programming for those who code as part of their work.

## Graphic tools

The share of respondents working in organisations using predominantly or entirely external graphic tools is increasing over time (from 45% in 2022 to 51% in 2023). Of these, in 2023 more than one in three uses solely external graphic tools. These shifts could point to the wish to seek out-of-the-box solutions that can provide quick and efficient results. This might be due to the limited resources to develop in-house tools or a preference for widely supported and continuously updated platforms. Largest organisations are those which predominantly enjoy in-house custom solutions when it comes to graphic tools.

## Tools

The tool distribution in 2023 is generally consistent with previous survey editions. Excel is the most commonly used tool, used by nearly three out of four of respondents. More than three out of five respondents use Google Sheets. Among data visualisation software, the gap between Flourish (32%) and Datawrapper (37%) remains constant. Several respondents who indicated "Other" have stated to use the following tools for data analysis or visualisation:



Power BI, RAWgraphs, Infogram, QGIS and various programmes belonging to the Adobe Suite, particularly Illustrator.

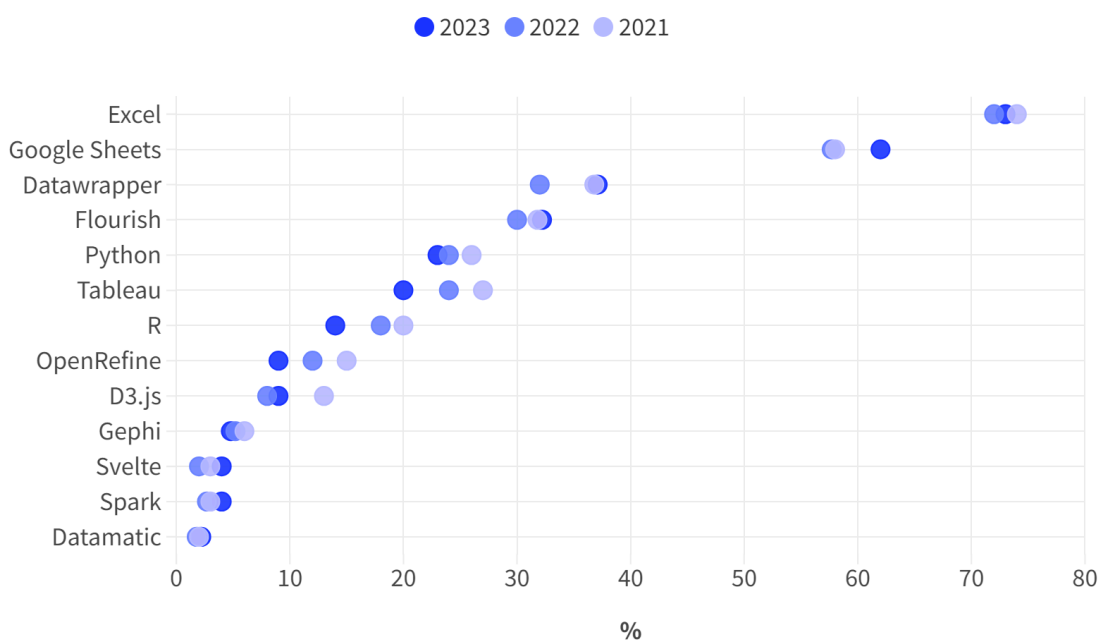


Figure 15: The usage of different tools for data analysis and visualisation.

There are stark national patterns in terms of tool usage, particularly for data visualisation software. While the percentage-point difference from the mean of Datawrapper users is +21 in the US and +18 in Germany, the Flourish equivalent is -5 and -17. Flourish is instead more commonly used in the UK (+21). Similarly, there are patterns in occupation breakdown too. Freelancers are less likely to use Datawrapper and Flourish compared to employees, editors, and educators. In terms of programming languages, Python is more commonly used among employees, while R is more commonly used by students and editors.



# Work Practices

## Type of data journalism

In 2017, Simon Rogers, Jonathan Schwabish, and Danielle Bowers published a report on the state of the field of data journalism. The findings showed data journalism outputs generally fell into three categories: investigative journalism, stories that explain data, and stories that are enriched by data.

As in 2021 and 2022, data journalism is predominantly focused on stories supported by data (61%), followed by investigative reporting (41%). This preference for narrative enrichment could be reflective of the broader audience's interest or the journalists' skill sets, or a combination of both. It is notable that with more experience in the field, journalists tend to diversify and conduct more investigative data journalism, which may require more time, experience and expertise.

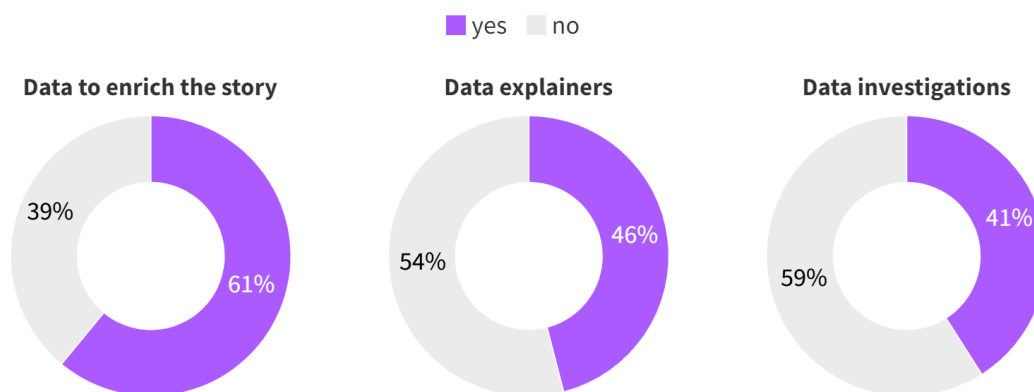


Figure 16: The types of data journalism by respondent performance of each type.



The gap between types of data journalism is wider for part-timers, who are more prone to use data to enrich the story and less prone to create data explainers than their full-time counterparts.

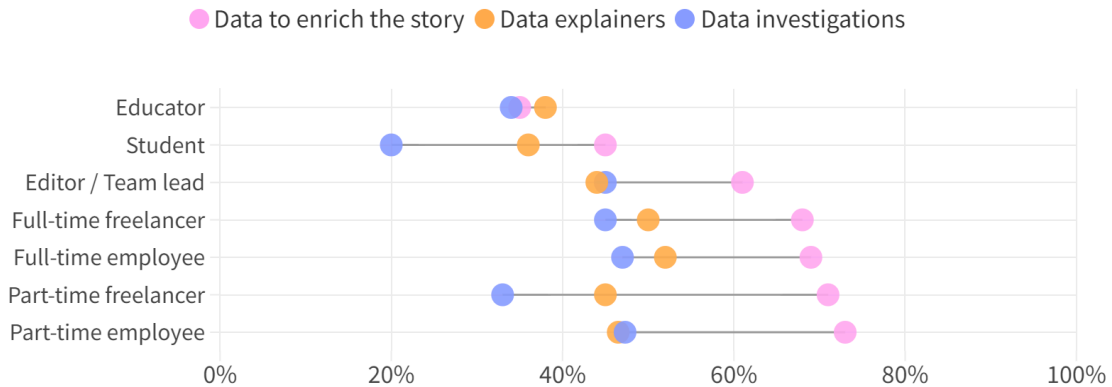


Figure 17: The types of data journalism, by occupation.

As years of experience in data journalism increase, so does carrying out any of the typologies listed, until this trend reaches a plateau around 3-5 years of experience. Such increase is particularly steep for data investigations, suggesting that versatility grows with time in the field.

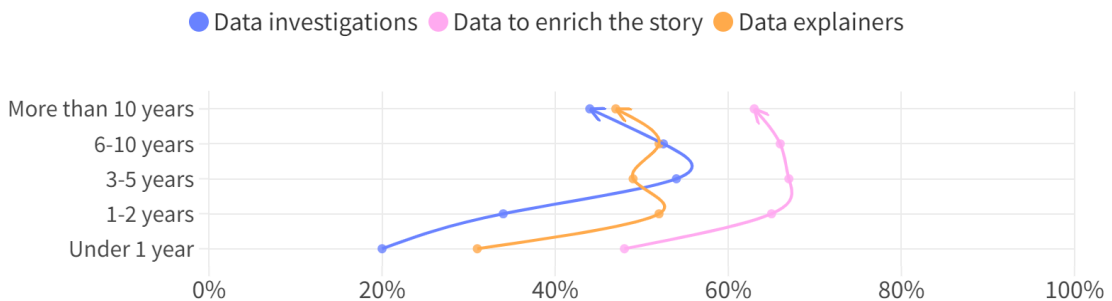


Figure 18: The types of data journalism, by years of experience.





## Work tasks

Data analysis is the most commonly performed task, carried out by nearly three out of four data journalists, highlighting the core importance of interpreting data for storytelling. In comparison, coding (19%), developing (9%), and web design (15%) are much more uncommon tasks. While three out of four analyse data, only just over half of respondents are also responsible for producing visualisations. We still find, like in 2022, that regardless of work tasks, as income increases so does the share of people who perform that work task.

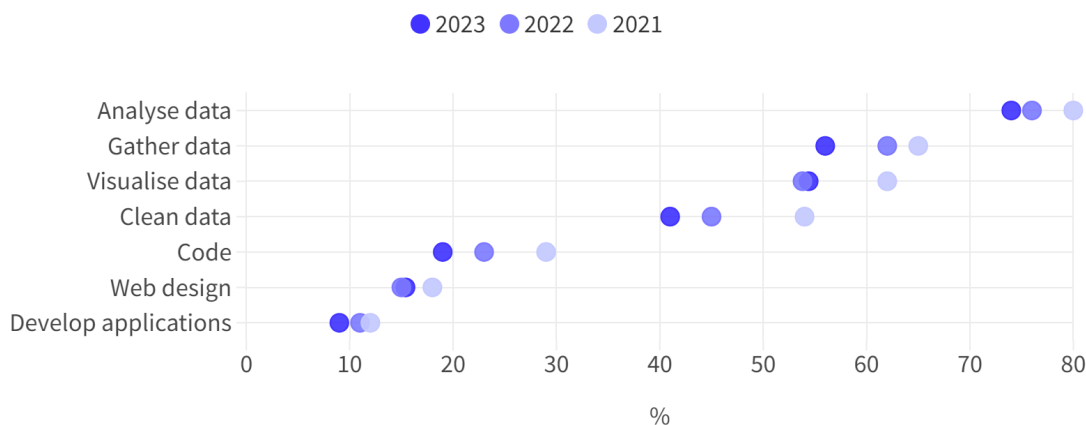


Figure 19: The distribution of work tasks by respondents.

Over time the distribution pattern of responses has remained constant. However, the share of respondents selecting each choice has generally decreased. This could be indicative of specialisation in one particular area, which is indeed suggested by the increase in fewer selections in this multiple choice question. The trend towards specialisation may be due to the increasing complexity and depth of skills required for each task, and it is particularly strong among women, who, like in the 2022 survey, chose fewer options in this multiple choice question.

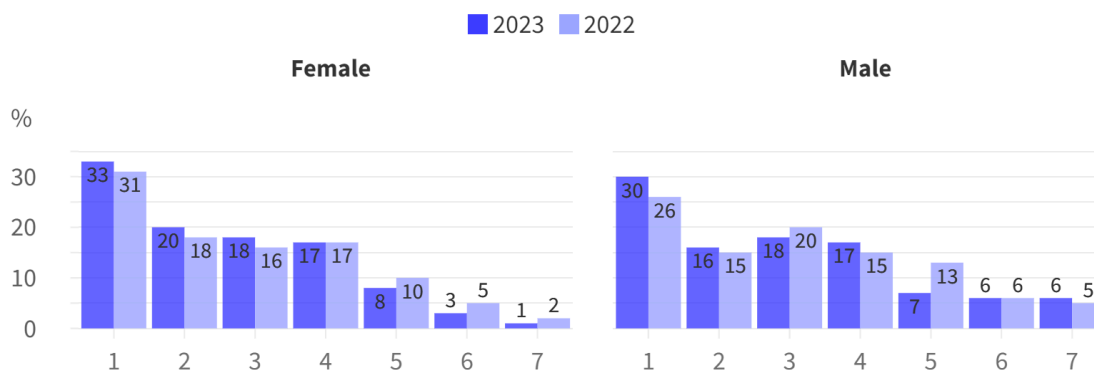


Figure 20: The distribution of number of work task selections, by gender.

## Scope

In terms of geographic scope, the largest group works at national media organisations (48%), where their work also addresses stories of national value or scale (50%). These figures are consistent with the results from the 2022 survey and indicate that national scale remains a significant focus for data journalism, or that many local newsrooms are yet to invest in data journalism. Only 22% work at local news organisations and only one in four produces local data journalism.

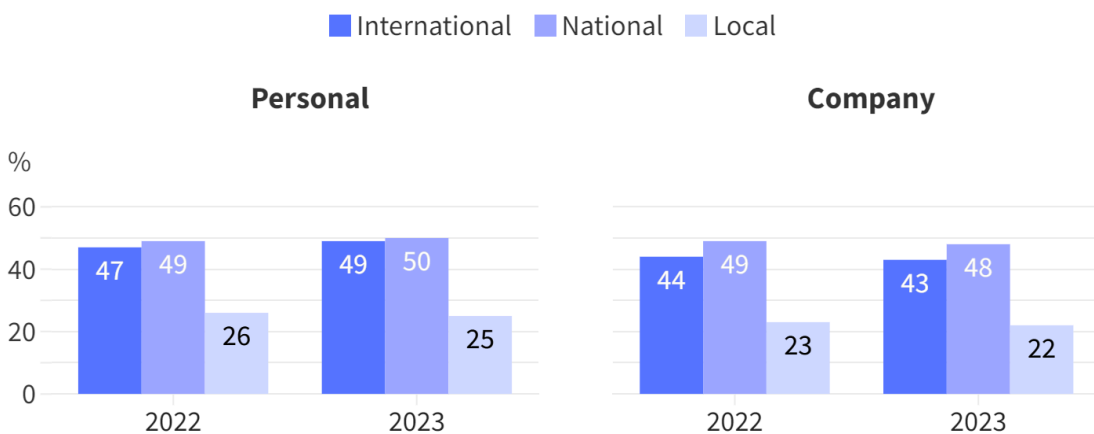


Figure 21: The scope distribution.



## Beat

Just over half of respondents cover Government and Politics in their data journalism, making it once again the top beat for the field. Environment and Weather / Climate reporting are steadily increasing since our first survey in 2021, with the former gaining six percentage points over the course of two years, and the latter eight. Weather / Climate is witnessing a sharp increase in coverage over time, as it was selected by 27% of respondents in 2021, 30% in 2022, and 35% in 2023. The Environment, which also has its highest increase in 2023, going from 44% in 2021, to 46% in 2022, and 50% in 2023, is the second most common beat, and it is now nearly on par with Government and Politics, reflecting the growing global concern over these issues and the strong relationship between data and climate reporting. The surge in climate and environment reporting highlights the role of data journalism in analysing and visualising complex datasets to shed light on environmental phenomena and the impacts of climate change. The Climate Journalism Awards and the 2023 News Impact Summit, organised by the EJC, responded to these shifts by incentivising journalists to adopt innovative techniques, driving greater awareness and action on environmental issues through data-driven journalism.

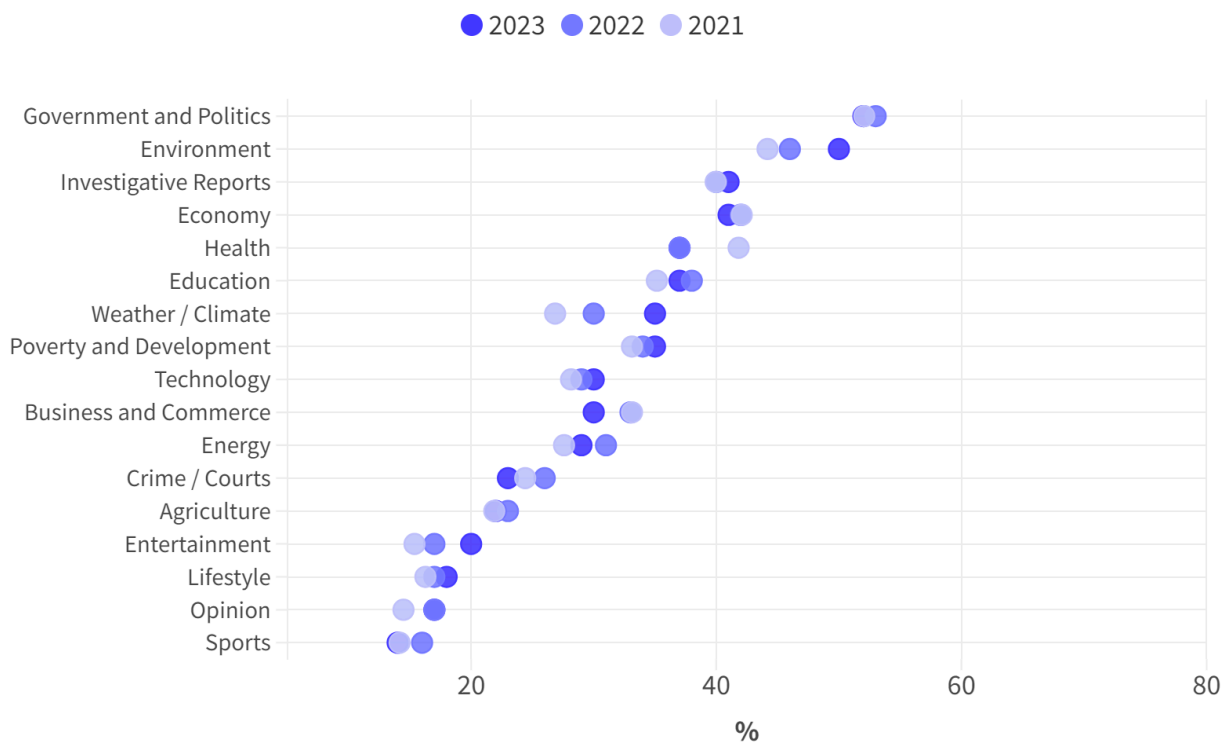


Figure 22: The beat distribution.



The share of journalists being specialised in more than one beat is on the rise. The median number of selections is five (standard deviation 3.75). One in four works on three and four different beats.

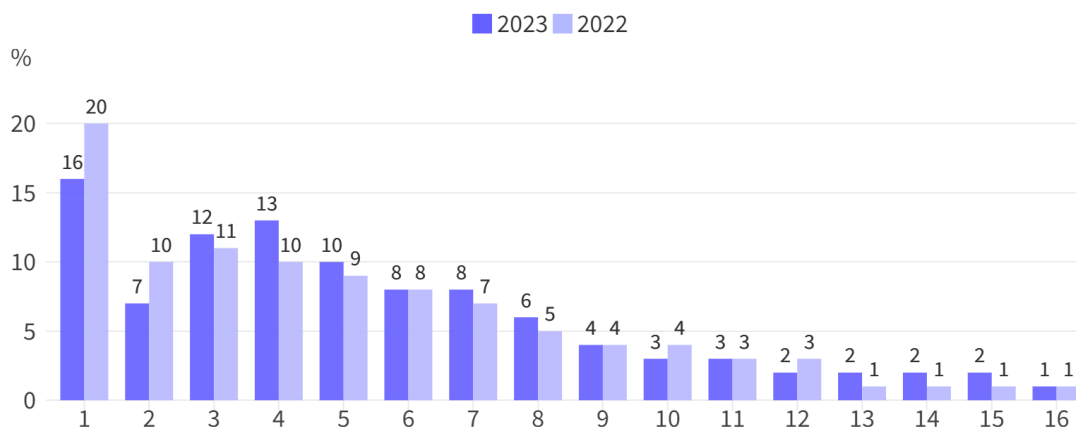


Figure 22: The distribution of the number of beat selections, by survey edition.

## Medium

Over half of data journalists work for an online-only digital outlet, making it the most common medium for data journalism and emphasising the digital shift in news consumption and the potential for interactive storytelling online. This is followed by print or broadcast media outlets with a digital site (38%). Data reporters working for TV or radio are just over one in ten, respectively.

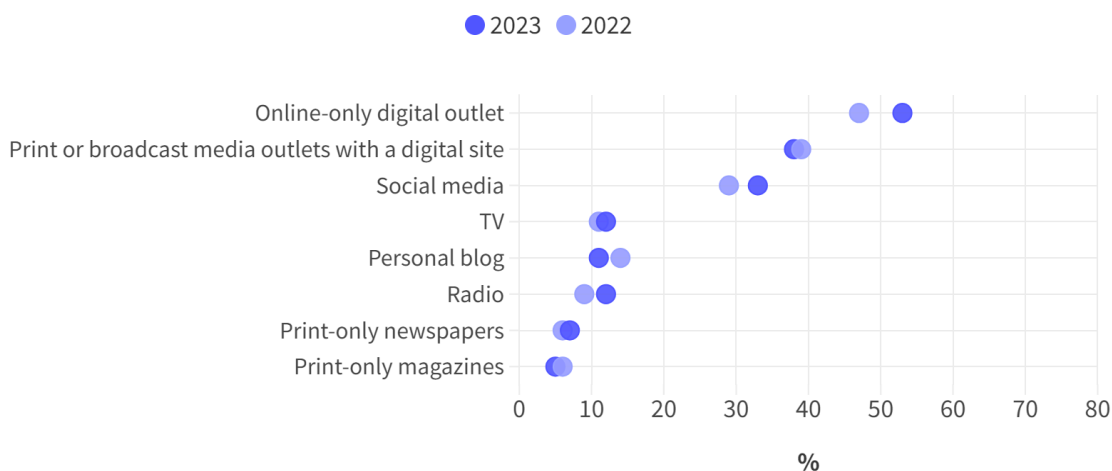


Figure 23: The distribution of medium types, by survey edition.

Like in 2022, over three quarters of respondents work for one medium in particular, or two. Oftentimes, combinations of different media types include social media alongside a primary legacy media format.

When it comes to the distribution of medium types by country, patterns remain overall consistent with the global distribution, in that online-only outlets and outlets with both a print or broadcast medium and additionally a digital site are the most common ones. Italy has distinctively low shares of respondents working for TV or radio.

Data journalists working in the TV media sector enjoy the highest salaries, followed by online media outlets. Print-only outlets are found at the opposite end of the spectrum earning the least amount in salaries. Since salaries are unadjusted by country, surely the distribution of medium types by country influences to a certain extent the patterns seen here.

## Data used

In 2023, public official governmental data remained the most common data type used, adopted by three out of four respondents. At the bottom of the distribution we find FOI-obtained data, used by one in five respondents globally. The widespread use of official governmental data signifies trust in and the necessity for authoritative sources, while the use of FOI-obtained data points to investigative journalism's need for transparency and uncovering information that



is not readily available. By country, FOI-obtained data is more commonly used in the UK (43%) and the US (40%). Census data is also most commonly used in the US, where 68% of respondents have reported having used it in 2023. Similarly, scraped data is most commonly used in the US (58%). On the other hand, social media data is most common in Russia (62%).

By profession, educators, followed by editors, and then by employees, work on average with the largest number of data types. Contracted reporters tend to work more with census data, governmental data, and FOI-obtained data than freelancers. Social media data and crowd-sourced data are instead evenly distributed across professions in data journalism.

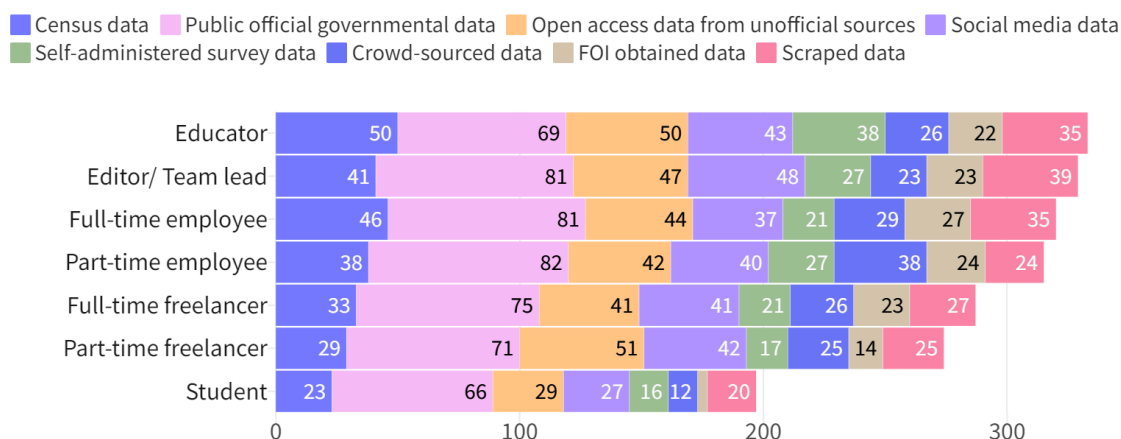


Figure 24: The distribution of data types used, by occupation.

## Dedicated data unit

One in four people surveyed works in a dedicated data unit. Dedicated data units are most commonly found in medium to large organisations, where around one in three of data journalists work in one. The presence of dedicated data units in larger organisations, but not as a norm, suggests that while there is a recognition of the importance of specialised data teams, many organisations may not have the resources to support them or may integrate data tasks across different teams. Regardless of company size, most dedicated data units (61%) have less



than five members, however larger teams of 6-15 individuals are on the rise compared to 2022.

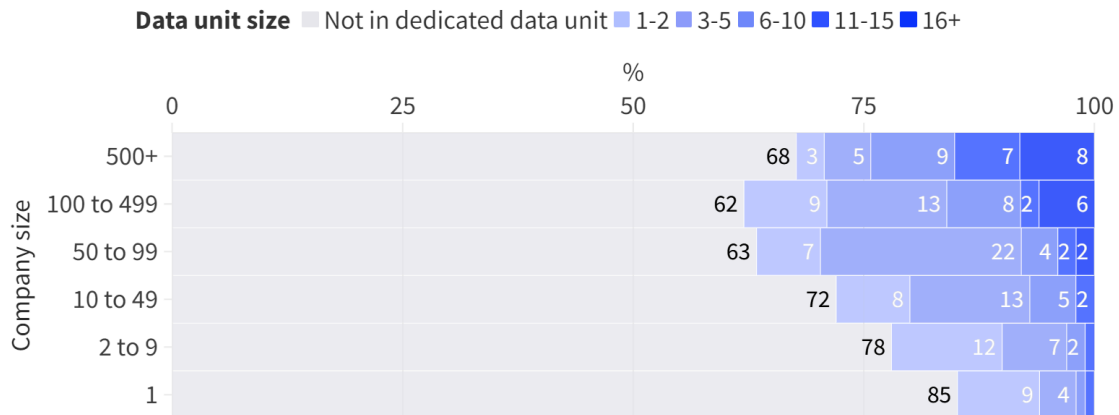


Figure 25: The distribution of data units by unit size, by company size.

## Projects

As the 2022 and 2023 survey results showed, most data journalism projects are completed within several weeks or months (51%), reflecting the time-intensive nature of this field. Projects completed in a day are very rare (9%), while one in five projects are completed in a week. Projects tend to be the fruit of collaboration with a small team of 2-5 individuals (45%), followed by solo work (32%).

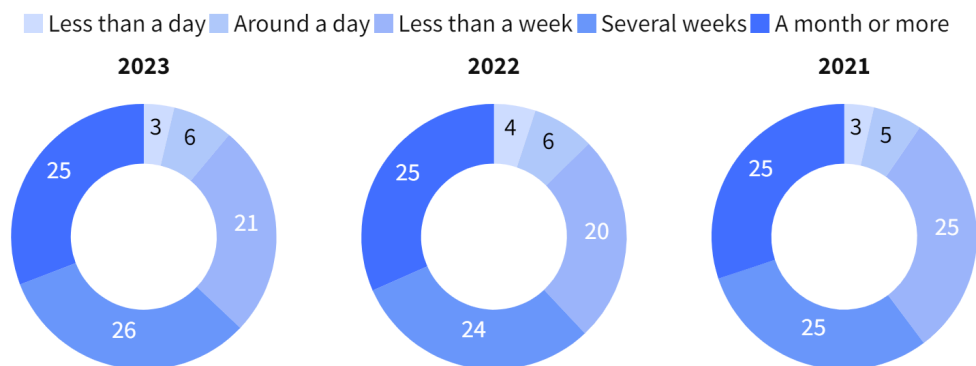


Figure 26: Project duration in the three survey editions.



## Collaborations

Collaboration projects between different organisations are relatively rare (29%) and temporally the pattern is stable with previous survey editions. 38% of collaborations are one-time occurrences leading to separate content production. In fact, one-time collaborations are more common than ongoing ones, and separate content production is much more frequent than sharing resources or production.

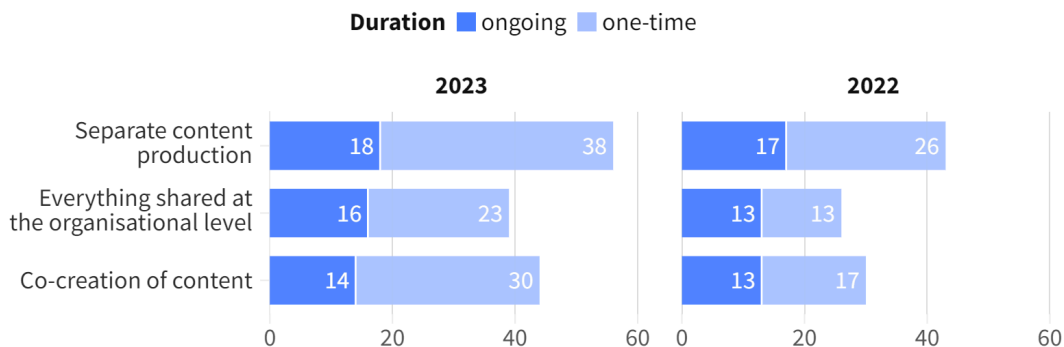


Figure 27: The distribution of different collaboration types.





# Challenges

## Challenges to producing more

Access to quality data remains the top hurdle for data journalists in 2023 (65%). Compared to 2022, this challenge has increased by eight percentage-points. The second most pressing issue is time pressure (48%), followed by a lack of adequate knowledge in data analysis (47%). Lack of interest from management is less of an issue in 2023 than 2022 and 2021 (a six percentage point decrease between the 2021 and 2023 survey editions). Lack of adequate software for data visualisation is also less of an issue in 2023 than 2022 and 2021 (a difference of 11 percentage points). These changes may suggest an evolving industry that is becoming more supportive of data journalism and its needs, for example through the development or improvement of data visualisation software.

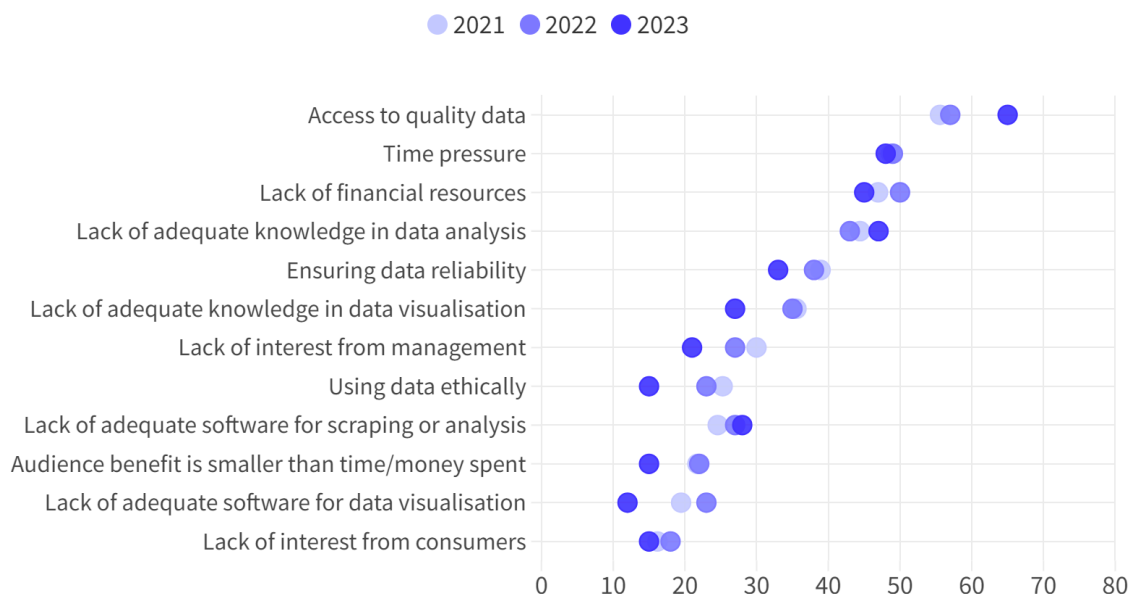
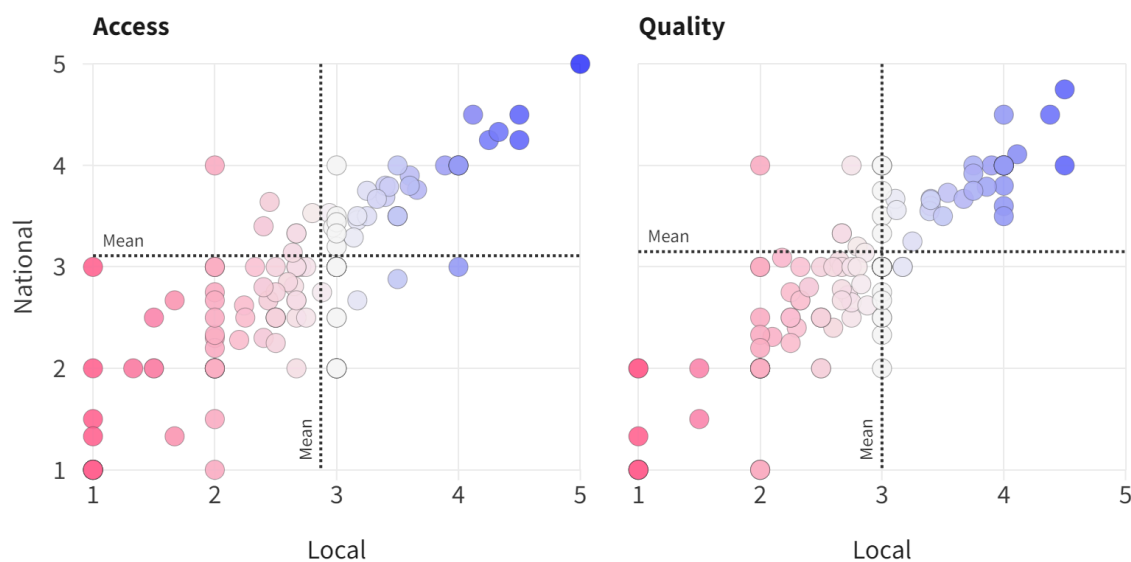


Figure 28: The different challenges encountered in the industry, by percentage of respondents.



## Data access and quality

The strong national variations previously noted when it comes to rating access and quality to local and national data in the country of work still persists. As in 2022, local data fares worse than national data, on both access and quality. Countries that tend to fare worse in one dimension, usually will either outperform or underperform in all dimensions. The persistent regional variations in access to and quality of local and national data suggests broader systemic issues related to data governance and transparency within different countries. They also underscore that depending on where you work you may also need to invest more time and resources in verifying and contextualising data to ensure its reliability and relevance. Addressing these challenges likely requires a blend of journalistic expertise, technological innovation, and advocacy efforts to ensure accurate and impactful reporting on critical issues.



Rating 1: very poor, 2: poor, 3: average, 4: good, 5: very good.

Figure 29: The rating of access and quality of local and national data, by country.



## Barriers to learning

The increasing barrier of financial resources (53%) suggests a need for more accessible and affordable education pathways in data journalism. While some obstacles, such as limited internship opportunities and entering the industry, have slightly decreased, challenges related to access to expert educators and practical experience persist. Efforts to mitigate these barriers should focus on expanding educational resources, promoting mentorship programmes, and advocating for inclusive practices within the industry. This underscores the importance of fostering a supportive environment that enables aspiring data journalists to acquire the necessary skills and knowledge to thrive in the field.

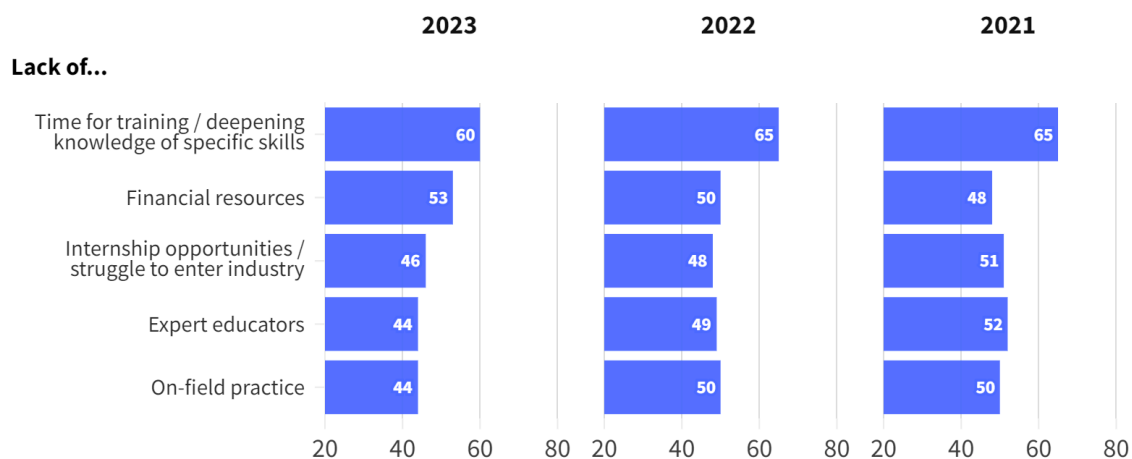


Figure 30: The barriers to learning faced in the industry.

## Value of data journalism

While the majority of respondents (70%) acknowledge the role of data journalism in enhancing the reliability of stories, there's a less unanimous view (29%) regarding its necessity solely due to the proliferation of data. However, data journalism is widely recognised for its ability to contextualise stories (61%), uncover unique newsworthy events (59%), and facilitate the discovery of



relevant stories (60%). Yet, there's a more divided opinion on its utility for event detection and monitoring (33%) and predictive analysis (24%). These findings emphasise the multifaceted value of data journalism in enriching storytelling and informing the public, while also highlighting areas where its impact may be perceived differently among practitioners and audiences alike.

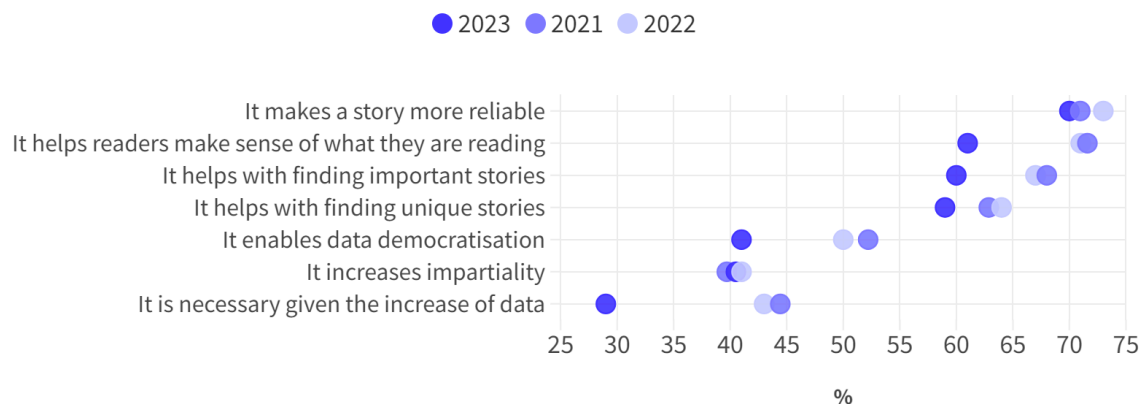


Figure 31: The self-reported value of data journalism.



# AI and OSINT

In 2023 we included a special module in the survey concerning usage of AI and OSINT. We found that just over one in three has used AI as part of their data journalism work. When it comes to Open-Source Intelligence (OSINT), the equivalent share is one in five.

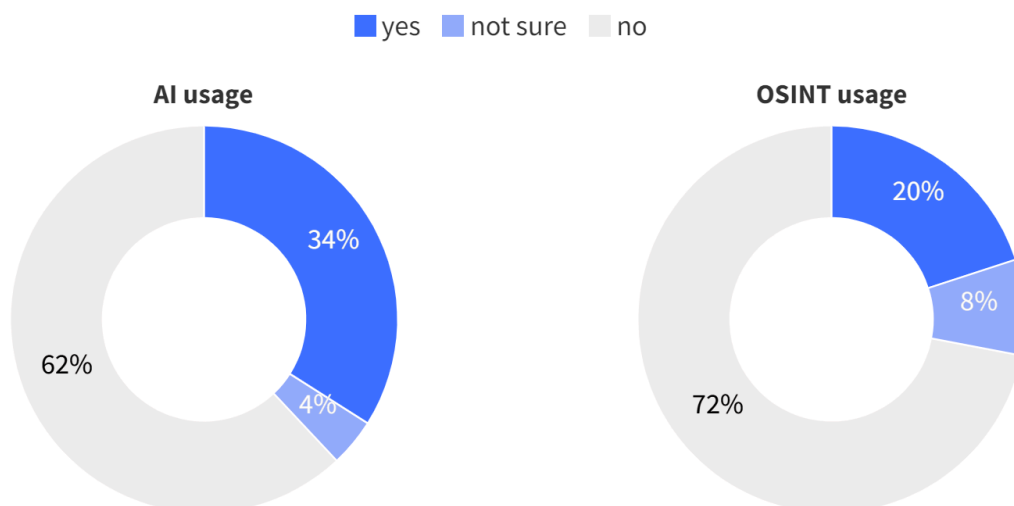


Figure 32: Usage of AI and OSINT for data journalism in 2023.

Across different types of newsrooms, AI and OSINT are used at differing rates. While AI is making its way in data journalism practices at similar levels whether the newsroom is local, national, or international in scope, there are large differences when it comes to OSINT, with 23% of respondents working in international newsrooms using OSINT, against 12% of those in local newsrooms. These patterns demonstrate the importance and value of AI for local newsrooms, with [particular benefits for reporters productivity](#).

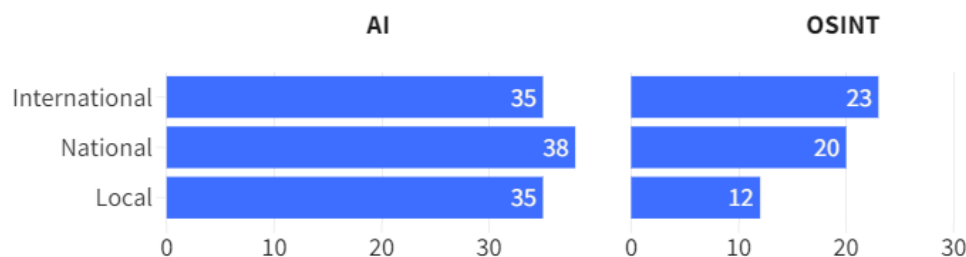


Figure 33: Usage of AI and OSINT by respondents across newsroom types.

## AI

Of those who have stated they have used AI, 45% have leveraged Artificial Intelligence to search for content. This is followed by content generation (38%) and text mining (37%). Much more rarely, AI is used to perform predictive modelling (8%). AI being used primarily for content searching suggests that data journalists are tapping into advanced algorithms to sift through large datasets, extract relevant information, or even identify patterns that might not be immediately obvious. This points to a trend where the role of AI is to augment the journalist's capabilities, allowing them to handle content navigation and generation more efficiently than traditional methods. This can include automated writing of reports on finance, sports, and elections where data is structured and predictable. Meanwhile, text mining can help uncover insights from unstructured data like social media posts, news articles, and research papers. There were also several respondents who used the "other" text field box to specify that they use AI as a programming assistant. One respondent explained that "I often ask ChatGPT to broadly help me think of two strategies in code (R or Python) with the use of certain packages to answer a question, and then build on that code and finetune it. I also ask ChatGPT about errors I've gotten and what to do with them", while another added that "GPT4 and GitHub Copilot have both increased my coding productivity significantly."

The data highlights distinct trends in AI task distribution across various scopes. Tasks such as Data cleaning / processing, Fact-checking and content verification, and Content generation are consistently prominent across all scopes. However, there is a noticeable concentration of tasks like Image and video analysis in international and national newsrooms, while Text mining and Data visualisation are more prevalent in local newsrooms.

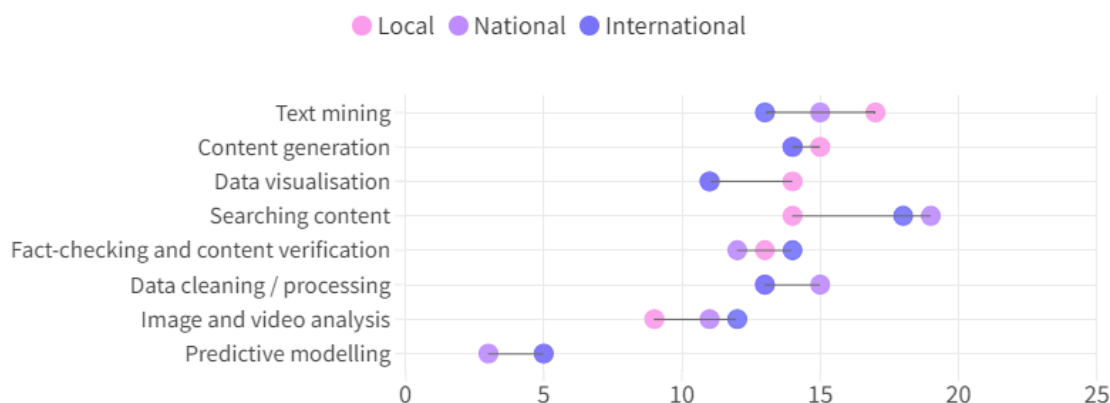


Figure 34: What AI is being used for in data journalism in 2023, by percentage of AI users across different types of newsrooms.

The type of AI tools most commonly used are generative AI tools, which have been used by 62% of AI users. The biggest challenge when it comes to using AI is a limited understanding of the tool or technology (56%). More than half of respondents are concerned about bias in and ethics of AI models. 41% stated they find a lack of dedicated time to experiment with the technology a barrier. The challenges expressed by journalists regarding AI, such as limited understanding and concerns about bias and ethics, are reflective of broader concerns in society about the implications of these technologies. It shows a need for better education on AI and a cautious approach to its application, ensuring that AI assists rather than dictates journalistic work. But there are also issues with unequal access to AI tools, with one respondent explaining that “Users from my country (Russia) are banned from using many of the AI tools due to sanctions and corporate policies”.

## OSINT

OSINT is used by nearly six out of 10 OSINT users to verify sources, images, or videos. This is closely followed by using OSINT to navigate maps, satellites, and location-based information (57%). The use of OSINT for verifying sources, images, and videos is increasingly becoming an essential component of journalism in the digital age, where misinformation and disinformation can spread rapidly online. By leveraging OSINT tools to authenticate content before publication, data journalists play a role in maintaining journalism’s credibility and trust with audiences. The use of OSINT tools to navigate and analyse maps,



satellite imagery, and location-based data may also indicate a journalistic trend towards more in-depth investigative work. Such tools enable journalists to verify claims and report on events from a distance, which is particularly valuable in hard-to-reach conflict zones or inaccessible regions for other reasons.

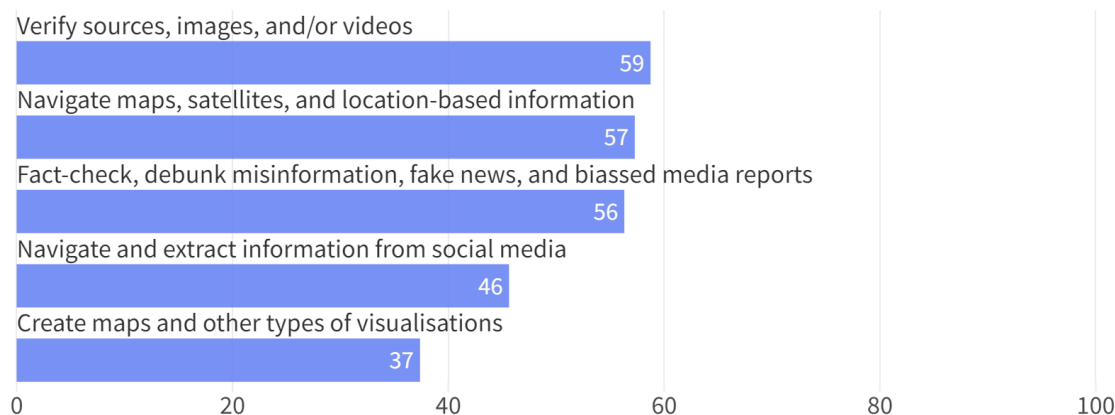


Figure 35: What OSINT is used for in data journalism in 2023, by percentage of OSINT users.

Most OSINT used comes in the form of public government data and reports (72%), while news media and internet media are used by around six out of 10 OSINT users. The prevalence of public government data and reports as sources for OSINT reflects the importance of transparency and official records in journalistic research. However, the use of news media and internet media from citizens points to the growing role of user-generated content in newsgathering and storytelling.



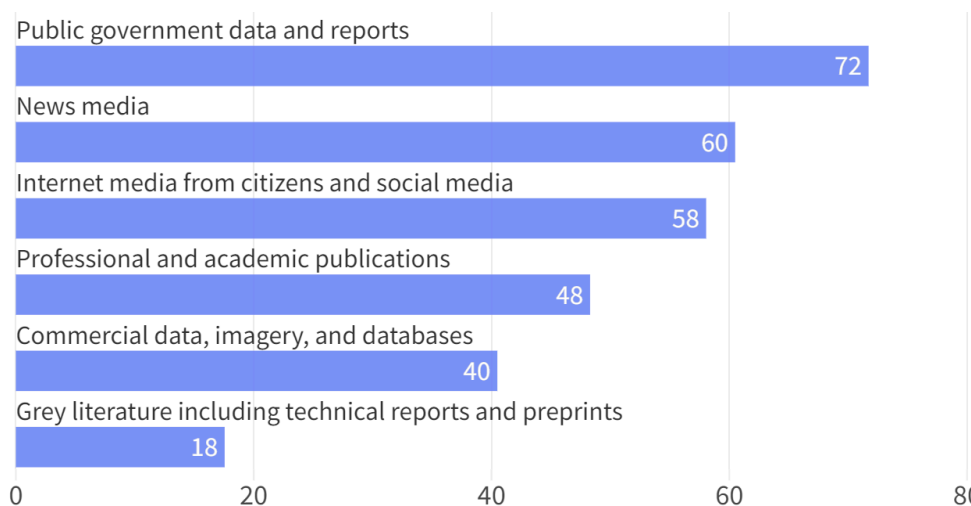


Figure 36: The type of OSINT used in data journalism in 2023, by percentage of OSINT users.

Open Source Intelligence users provided a comprehensive overview of the challenges associated with OSINT. One respondent voiced frustration over lack of formal educational resources, stating, "There's no formal training or school to learn OSINT skills for journalists, leaving many of us to figure it out on our own." Evaluating the quality of social media data appeared as another challenge, with another participant stressing that, "Better skills are needed to understand how to do this effectively and in a timely manner." These challenges imply, as respondents suggested, that "OSINT is very time intensive". This is particularly the case for freelancers who lack a supportive network for data validation, one respondent indicated. Access to essential verification tools also posed a significant barrier, as noted by a participant who remarked, "Tools to verify information or fact-check data are not open source and often require expensive subscriptions or are locked behind paywalls or private institutions who hold the license" Concerns were also raised regarding the removal of user-generated content by tech companies, with one respondent stating, "Many big tech companies are removing sensitive data because it violates platform policies. But when this data is removed, it means it disappears for documentary evidence purposes."

Despite these challenges, respondents shared valuable insights into strategies for resource acquisition and skill enhancement. One suggestion included subscribing to OSINT resources like newsletters and following hashtags on social media. Another respondent learned by reading "the methodologies from



investigations published by reputable newsrooms”. Several institutions were mentioned as providers of useful case studies, including Bellingcat, GIJN, ICFJ, and EJC.

Ethical considerations also played a role in respondents' perspectives, with an emphasis on data security and privacy. As one respondent put it, "Data security is key for privacy and digital security reasons when it comes to working with OSINT data." Another respondent talked about the balance between public interest and the right to privacy, stating, "Sometimes sensitive data is shared in a story, but this is often when the importance of the story is heavily in the public interest. Sharing sensitive data should be proportionate and justified. "

In conclusion, the findings from these survey modules underscore the growing significance of Artificial Intelligence and Open Source Intelligence in data journalism practices. AI primarily aids in content searching, generation, and text mining, enhancing journalists' capabilities. Challenges include limited understanding, bias concerns, and unequal access to AI tools, emphasizing the need for education and cautious implementation.

OSINT plays a crucial role in combating online misinformation, aiding journalists in source verification and investigative work. Its significance in addressing misinformation will likely grow in the context of upcoming elections and the reporting of ongoing conflicts like those in Ukraine and Gaza. By utilising tools such as satellite data and social media analysis, journalists can access hard-to-reach locations and provide accurate reporting. The steps needed and methods undertaken to generate a piece of factual reporting are used to promote the validity of the coverage and trustworthiness of the media brand, as BBC Verify attempts to do.

Finally, we already see that the marriage of AI and OSINT creates a powerful toolkit for data reporting, as indicated by one respondent: "AI algorithms assist me in sifting through large datasets and identifying patterns, trends, or anomalies that may not be immediately apparent through traditional methods. For example, I use machine learning models to automate the categorization of vast amounts of information collected during OSINT processes. These models can help classify data into relevant topics, themes, or sentiment, enabling me to quickly distill insights and focus on key areas of interest."



# Conclusion

In 2023, the world of data journalism saw some interesting changes while keeping some familiar aspects. As demographics shift, we witness a stride towards gender inclusivity and the influx of fresh, young talent. Yet, traditional occupational structures persist, offering both stability and flexibility in work arrangements, and underscoring the resilience of established media frameworks.

Skill development remains a focal point for the data journalism industry, with a strong demand for bolstering technical proficiencies, particularly in data analysis and visualisation, to meet the evolving demands of modern storytelling effectively.

Intra-organisational collaboration is key in data journalism, with small teams working together on projects that typically take a long time to come to life, underscoring the interdisciplinary nature and time-intensive efforts inherent in this field. Dedicated data units remain niche and typically situated within medium to large organisations.

Challenges persist, from data access hurdles to time constraints and skill gaps, echoing the need for enhanced data accessibility, efficient workflow management, and ongoing skill enhancement initiatives. Yet, concerns over management buy-in and software support are becoming less of an issue, reflecting an evolving organisational landscape increasingly supportive of data-driven endeavours.

Technological advancements, albeit modest in adoption, signify that AI and OSINT are gradually finding their place in content search and verification. However, ethical considerations and challenges call for greater education and ethical frameworks to navigate the complexities of these tools effectively.



# Thank you

A warm thank you to all of the people involved in data journalism who took the time to participate in our survey.

A heartfelt thank you to all those who helped us craft and polish our survey before launch.

And to the supporters of Datajournalism.com, for offering tools, goodies, and insights.

Without the support of other organisations, whether financial or in-kind, the European Journalism Centre would not be able to support the data journalism community. We welcome conversations with those who are appreciative of our work and want to support initiatives such as this report.