



# The State of Data Journalism 2021

---

The full report of results from the biggest data journalism survey of the year.  
Brought to you by [DataJournalism.com](https://DataJournalism.com)



# Table of Contents

Our Mission	2
Methodology	3
Key Takeaways	4
Demographics	5
Employment	8
Skills and Tools	11
Work Practices	14
Challenges	21
Pandemic and Work	24

This report is a version for print of our State of Data Journalism 2021 results. The plots were designed with interactivity and web in mind, so to enjoy the full potential of the charts, please navigate them on our results site.



# Our Mission

Data journalism is now an established part of the media ecosystem, with many newsrooms having a dedicated data team and others looking to create one. The formalisation of the field is less than a decade old, and the practices, skill sets, and technologies used are rapidly evolving, as discussed in our latest edition of the Data Journalism Handbook.

We think the establishment status of the field means that data journalism deserves to be studied, mapped, and taken seriously. But we also see that its rapid evolution indicates the need for continuous snapshots to understand how data journalism is conducted and how it changes over time.

Another exciting reason to measure the field is exemplified by the imaginative, collaborative, problem-solving nature of data journalism. Data journalism is made by individuals who blend together different data sources, analysis tools, and visualisation, to create powerful storytelling. What is possible then, and how are things being done, we ask?

Yet we also acknowledge the challenge to upskill beyond the realms of journalism. This includes learning statistics, data visualisation and programming. And further to that, to keep up with the pace of evolving work practices. As time progresses, new tools get adopted, and some see the dawn of day. Team struc-

tures change, and new job opportunities arise. All the while, data journalists are affected by the same struggles of other media players: shrinking resources, time scarcity, and waning public trust in journalism.

Within this complex landscape, the COVID-19 pandemic has brought new challenges to data journalists, but it also put them in the spotlight thanks to a wider audience. In the survey, we ask what impact the pandemic has had on data journalists' work practices.

The reflections presented were the driving factor that led us to launch the State of Data Journalism 2021. At present, the field is lacking a regular and systematic approach that can help us make sense of the role, modus operandi, and industry composition of data journalism. Previous efforts include a 2017 survey by Heravi and Lorenz<sup>1</sup> and a Google News Lab report<sup>2</sup> from the same year.

These studies generated useful and unique insights. Yet much has happened since 2017. We build upon the learned lessons from these authors and create a survey that poses new, relevant questions, helping us understand the field today, in 2021.



# Methodology

The State of Data Journalism 2021 Survey comprised a total of 63 questions and had 1594 respondents, of which 1285 were used for the analysis. The survey was organised in several sections, ranging from demographics to work characteristics, challenges and opportunities, and the impact of the COVID-19 pandemic on data journalism. It was available in four languages: English, Italian, Arabic, and Spanish.

## 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, as well as freelancers; data editors and team leads; trainers, faculty members, educators; and students. When it came to a sampling strategy, we faced the similar issue of Heravi and Lorenz<sup>1</sup> of not having any information on the global data journalism population parameters. Discarding the possibility to draw a random or a representative sample, we then, similar to Heravi and Lorenz, followed the approach of trying to reach as widely as we could through a variety of influential channels. Settling for a margin of error of 0.05 and a

confidence level of 95%, we estimated that we wanted to obtain a minimum of 350 respondents. We believe our analysis sample size of 1285 provides us with stable estimates. As an additional check, we used Diffbot's Knowledge Graph to check the properties of what they gather as the current data journalism population worldwide, against those of our sample. We obtained a very similar gender distribution, with theirs being 63% male and 37% female.

## Outreach strategy and incentives

The survey was open between November 8 and December 31, 2021. Participation was encouraged through various communications channels, to minimise bias obtained by only targeting one online community. We used direct mailing, social media promotion, and asked the DataJournalism.com and European Journalism Centre network for help in spreading the word. As a means to incentivise participation, while thanking for partaking, we offered a selection of rewards, which have now been distributed. The prize drawing included one trip to the International Journalism Festival 2022 in Perugia, Italy, Amazon vouchers, and digital goodies.



## Survey logic

To minimise survey length while maximising survey inclusion, questions targeting a specific subgroup were only shown to those respondents, but 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 data journalistic work from time to time). The survey included a mixture of mandatory and non-mandatory questions.

## Metadata

The median time spent on the survey was 16 minutes. Eighty-four percent of respondents took the survey in English, followed by 7% who took it in Italian, 6% in Spanish, and 3% in Arabic.

## Data Cleaning

For the purpose of the analysis, only complete answers were considered. The definition of complete is that a person has finished the survey, meaning he/she/they clicked on “Submit” at the end of the questionnaire.

We found 13 duplicate names, an indication that some people filled the survey twice. Whenever the name was the same but the email address was different, we looked for additional information in the filled questionnaire (such as company, job role, and skills) to determine if it was the same person or not. Instead, whenever the email address was the same we directly determined it was the same person who had taken our survey twice. Out of 13 duplicate names, we removed 11 duplicate questionnaires. We randomly selected which to keep for each person.



# Key Takeaways

Explore here below some of the key takeaways from our various survey sections.

## Demographics

There are more men (59%) than women (39%). Women are on average younger than men. 9 in 10 data journalists have completed higher education studies.

## Occupation

There are few part-time employed data journalists, whereas this is evenly distributed in freelancing. Women are overrepresented in freelancing and student positions, and underrepresented in editorial/team lead positions.

## Projects

Most stories are produced over several weeks or over a month. It is uncommon for data journalists to create a story within a day. 27% of data journalists have been part of a collaborative project in the last year.

## Dedicated data unit

22% of data journalists work in a dedicated data unit. Most data units are between 3-5 individuals, and otherwise 1-2.

## Experience and skills

Around 63% have less than 5 years of experience in data journalism. Data journalists feel much more advanced at journalism (60%) than data analysis (21%) or visualisation (17%). Just over 1 in 4 uses programming for work.

## Data

Newsrooms are those who leverage on FOI data the most, while students heavily rely on publicly available data. More income means working with more data types, particularly with scraped and self-administered survey data.

## Challenges

The 3 main challenges for data journalists are of institutional type: lack of quality data and financial resources, and time pressure. Lack of interest from consumers and audience benefit are the least commonly identified challenges.

## COVID-19 pandemic

1 in 4 has become a data journalist due to the pandemic, and 56% have worked on a COVID-19 related project. They see the pandemic as having strengthened the field and improved audience data literacy. Time pressure is where they have been mostly negatively impacted.



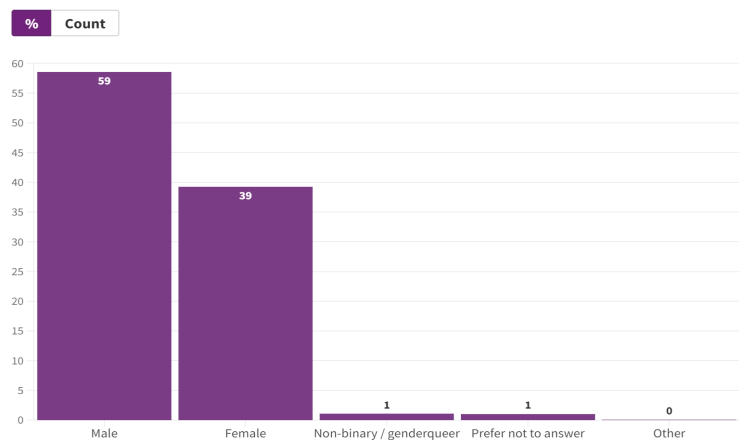
# Demographics

## Gender

Out of the 1274 complete responses, 59% of respondents identify as men, 39% as women, 1% as non-binary/genderqueer and 1% preferred not to answer. We also had 1 respondent identifying as other. Our results confirm previous studies showing a higher proportion of men against women in data journalism, although our research is the first that we know of that includes non-binary/genderqueer in the breakdown. For example, Heravi and Lorenz's 2017 study<sup>1</sup> had 57.5% of respondents identifying as male and 42.5% as female.

### What is your gender identity?

Number of respondents: 1274

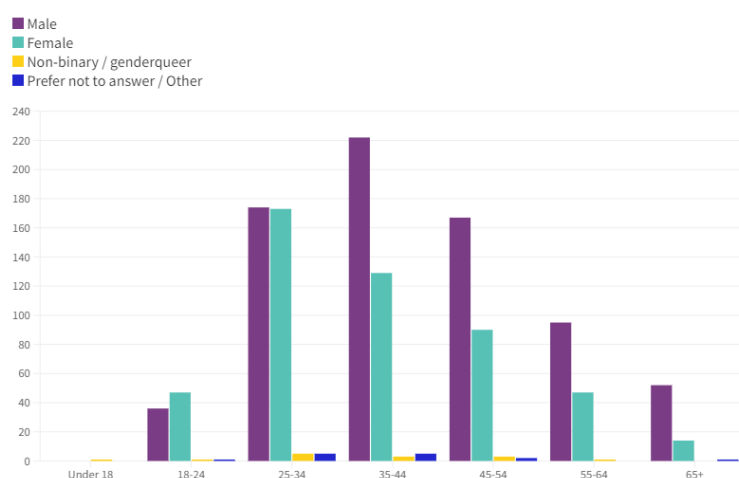


## Age

The data journalism community tends to be relatively young, with most people being between the ages of 25 and 44, corresponding to 56% of the sample. When we split the distribution by gender, we see that women in data journalism are both younger than men, with 35% of them aged between 25-34 years old. The most numerous age group for men is instead between 35-44 (30% of all men). The mode for genderqueer/non-binary individuals is also represented by the 35-44 age range.

### What is your age range?

Number of respondents: 1274





## Country of work

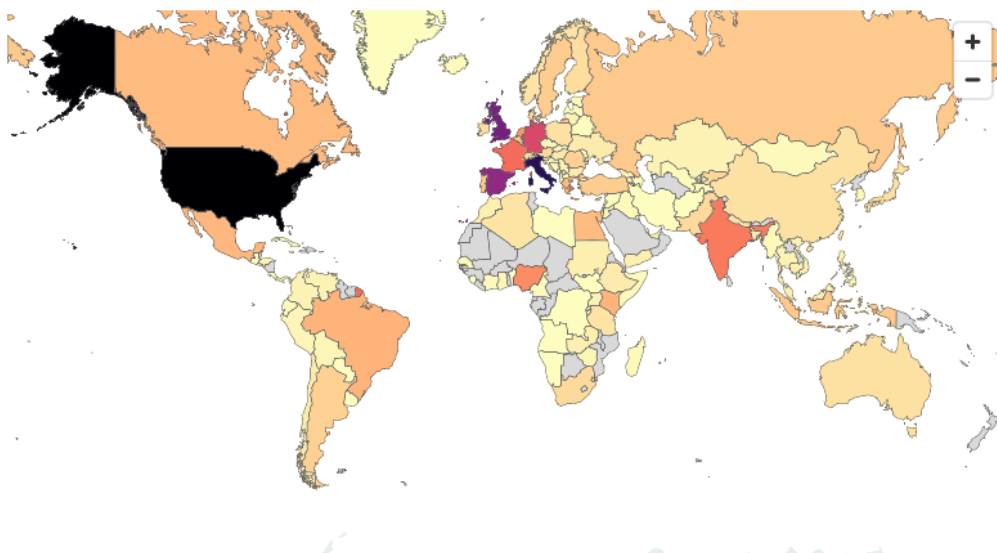
The respondents' geographic distribution in terms of country of work sees only Western countries in the top five, altogether representing 37% of the sample. The results show the United States (11%) as the top country followed by Italy (9%), the UK (7%), Spain (6%), and Germany (4%). Yet, in the top 10, we see countries from different regions such as India (3%) and Nigeria (3%). For Latin America, Brazil was the country with the highest response rate (2%). Our geographic distribution differs quite significantly from that of Heravi and Lorenz1, where they had 16% of participants coming from countries outside Europe and North America. In our case, 38% of participants come from a country outside the EU and North America.

In terms of the gender divide, the smaller the sample the more variability we see from the overall gender distribution. In the top countries, in terms of the number of respondents, Germany presents itself as the most inclusive and egalitarian country, with 45% of women and 3.64% of non-binary/genderqueer. Spain is also more evenly divided between men (55%) and women (45%) than the average. With 16 respondents in each, Portugal and Russia are the two countries in the top 20 where women are a majority (69% respectively).

### In what country do you work or study?

Number of respondents: 1274

% 0 10.52







## Education

Data journalists tend to be highly educated, with 91% having completed a higher education course or degree. Of those, about 47% has obtained a Master's degree, showing that many people choose to specialise after completing their Bachelor's degree. While university degrees show an even distribution of men and women overall, professional degrees, technical institutes and lower education are disproportionately male, showing that more women than men pursue higher education studies.

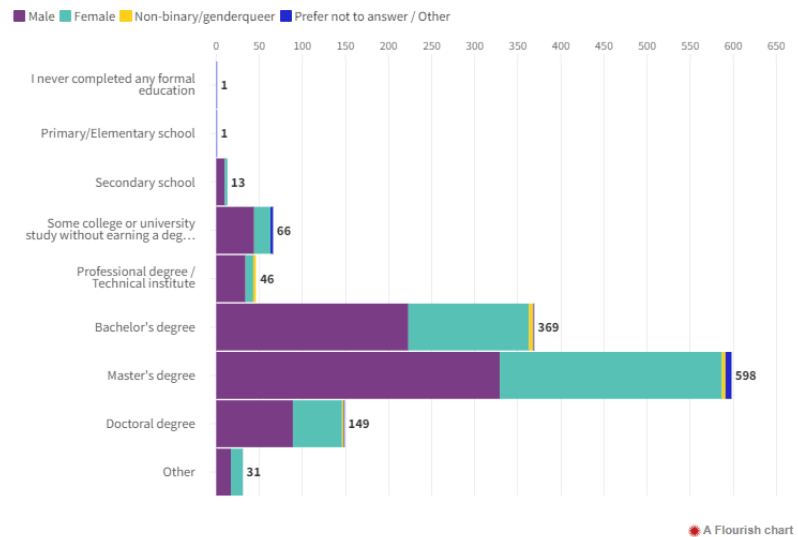
We asked those who completed higher education degrees to tell us the disciplines they chose for their studies. Around half of respondents (52%) have a degree in Arts and Humanities (which includes journalism), followed by the Social Sciences (20%), and then technical disciplines such as Computer Science, Mathematics, and Statistics.

We wanted to see the breakdown of disciplines by degree type but were challenged by the fact that, wherever a respondent picked more than one discipline, we could not link each selected discipline to the correct degree type. To account for this, here below we only show the breakdown for those who only selected one discipline (944 of 1161).

Looking at how they divide across degree types, we see that most Arts and Humanities degrees are obtained at the Bachelor's level (29%), showing that people chose thereafter to specialise in other disciplines. Survey participants with more advanced degrees were more likely to specialise in Social Sciences, Formal Sciences and Natural Sciences.

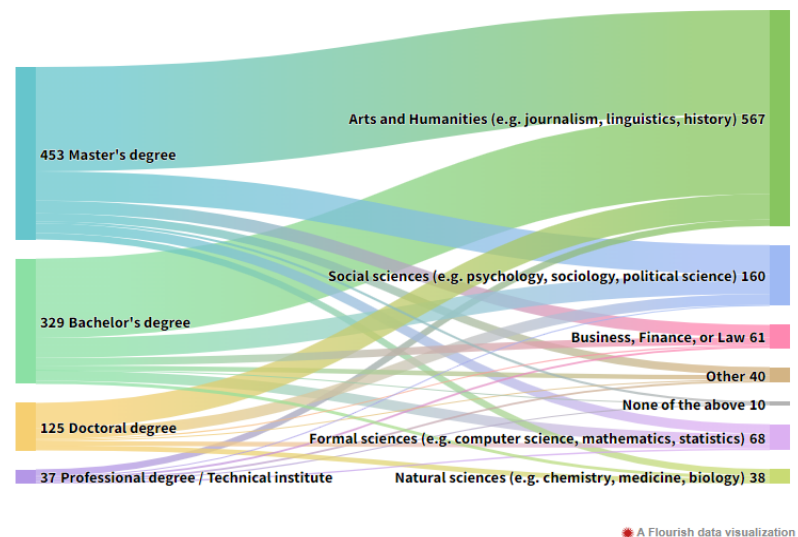
### Which of the following describes the highest level of formal education that you have completed?

Number of respondents: 1274



### In what discipline/s did you receive your degree/s? (Select all that apply)

Filtered by those who only selected one option, to allow for the correct linkage between degree type and discipline. Number of respondents shown in plot = 944





# Employment

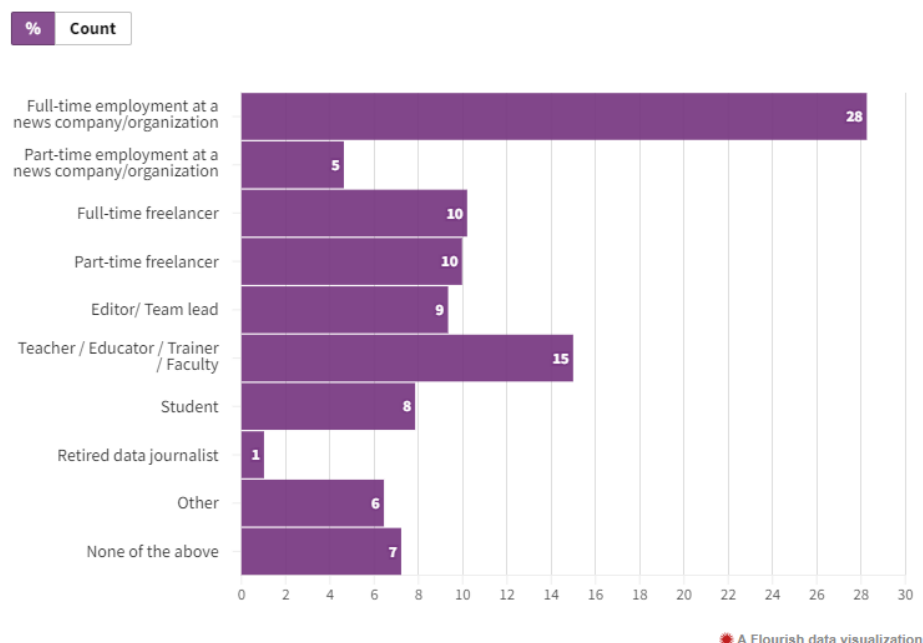
## Occupation

We asked our respondents to describe their primary connection to data journalism, with the aim to obtain a clear picture of the composition of the data journalism ecosystem. As a primary occupation, by far the largest group is of full-time employees at news companies or organisations (28%). While for every part-time employed data journalist there are 5-6 full-time employed ones, the share of full-timers and part-timers in freelancing is almost the same (10% and 10%). Educators, teachers, trainers, and faculty members are the overall second largest category (15%) among the survey participants. Meanwhile, students made up 8% of those who took the survey, an indication showing data journalism is a growing field.

As we break down occupation by gender and age, we once again see that women are over-represented in the student population, which is predominantly a young one. Women are also a majority when it comes to full-time freelancing, and are generally overrepresented in free-lancing roles. The youngest respondents (18-24) tend to be students or work as part-timers. Not surprisingly, older respondents predominantly tend to be in education-related positions or retirement. Looking at the 10 countries with the largest number of respondents, we see that Nigeria has the highest internal share of employees, whereas Italy, Germany, and the Netherlands have even shares of freelancers and employees. Brazil instead stands out for its share of educators/trainers (32%).

### Which best describes your connection to data journalism in 2021?

Number of respondents: 1273





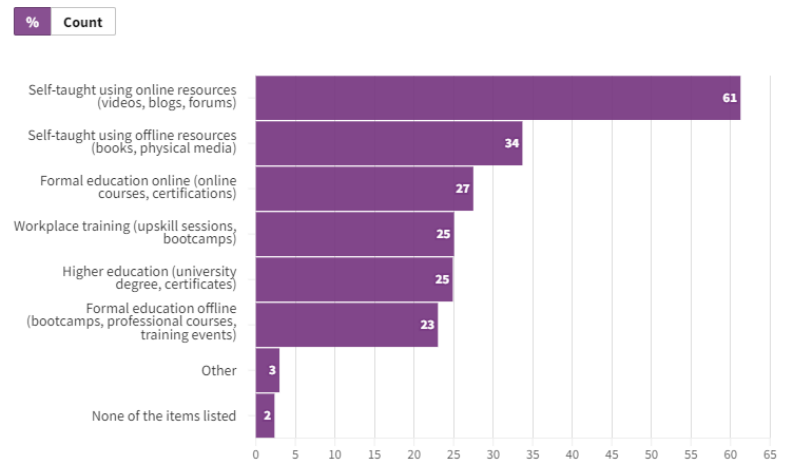
## Learning Data Journalism

We know that data journalists tend to be highly educated, but what education path do they tend to take when learning data journalism specifically? In our multiple-choice question, 70% indicated that they are self-taught using either online or offline resources, if not both.

Formal education online (27%) is a more popular learning place than higher education or workplace training (25% each), and bootcamps or professional courses (23%).

Workplace training occurs relatively more within those professions, such as full and part-time employment, or editor/team leads.

### Which best describes your data journalism formation? (Select all that apply)



Because this question was a multiple choice one, the number of entries exceeds the number of respondents, which leads to the accumulated percentages to be over 100%.  
Entries n = 2556.  
Respondents n = 1273.

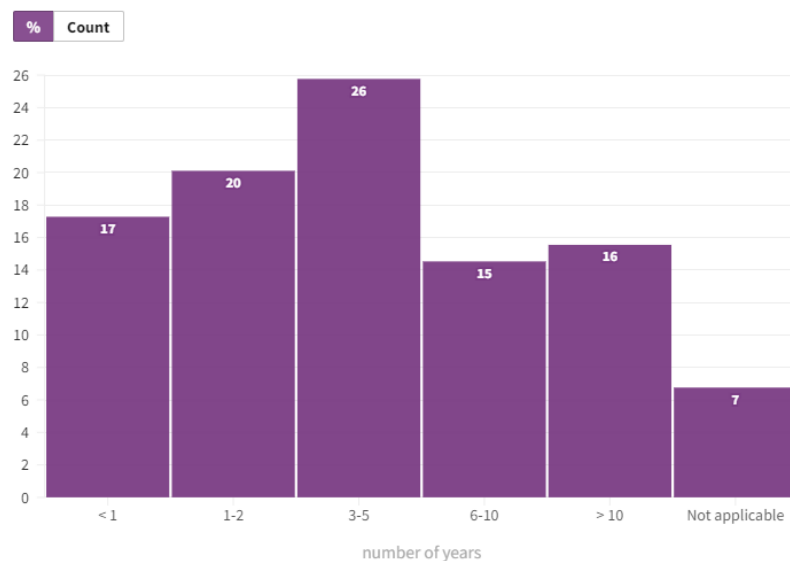
A Flourish data visualization

## Years of experience

Around 63% of data journalists have less than 5 years of experience in the profession, a reminder that many are in the early stages of their data journalism career. Full-timers, editors, and educators have by far more experience, in the number of years than the other occupation categories, with the exception of retired data journalists, who have been in the field the longest.

### How many years of experience do you have in data journalism?

Number of respondents: 1273



A Flourish data vis

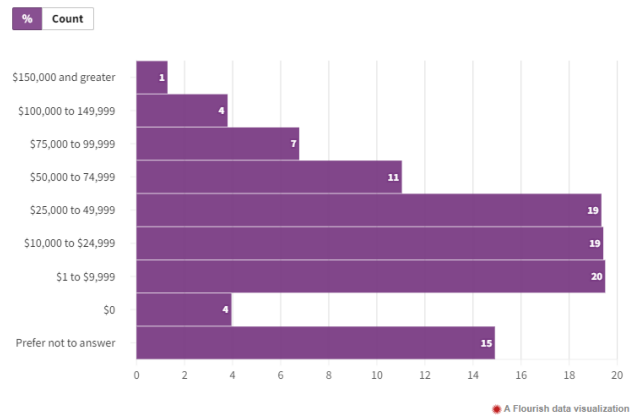


## Income

Of the 1241 who answered our income question, just under 60% earns between \$1 and \$49,999. By gender, men tend to earn more than women. Salaries increase with years of experience, age, and company size. Among professions, educators and trainers stand out in terms of quality of salaries.

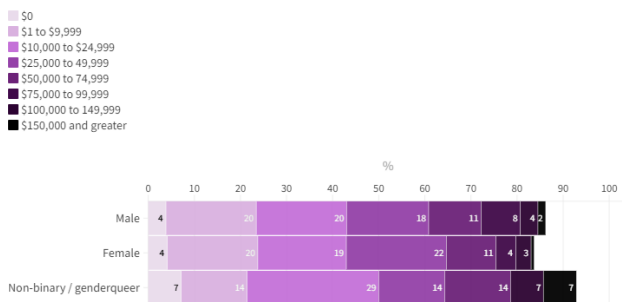
### Which of these describes your personal income last year? (Annual salary in USD)

Number of respondents: 1241



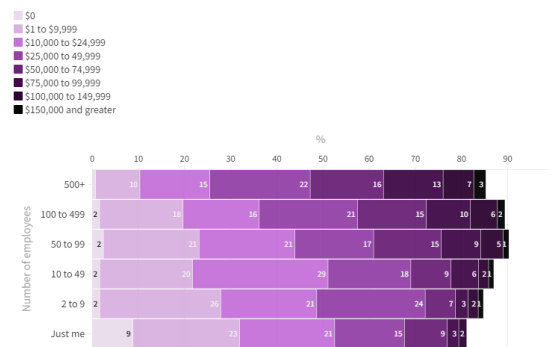
### Which of these describes your personal income last year? (Annual salary in USD)

Number of respondents in the plot: 1047



### Which of these describes your personal income last year? (Annual salary in USD)

Number of respondents in the plot: 1229

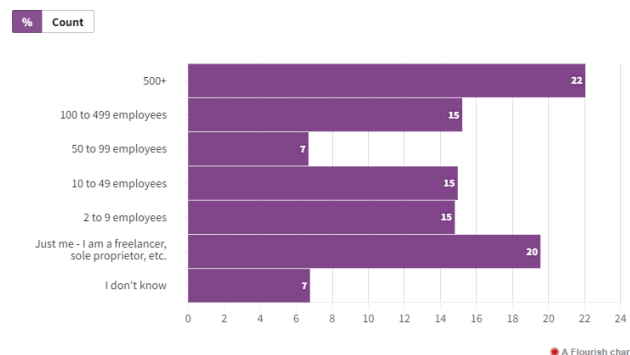


## Company size

It looks like the majority of data journalists like to join large organisations (22%), or else work on their own (20%), with small-medium to medium-large companies hosting the remainder (15% each in companies of 2-9, 10 to 49, and 100 to 499 employees, and 7% in companies of 50 to 99 employees).

### Approximately how many people work in your company/organisation?

Number of respondents: 1243





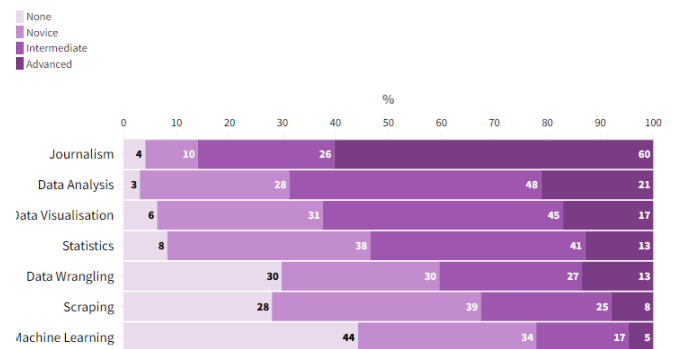
# Skills and Tools

## Skill level

How data journalists rate themselves in terms of skill level varies. Journalism is the area where data journalists feel they most excel, with 60% considering themselves to be advanced in this area. As for data analysis, 21% of those surveyed describe themselves as advanced.

When it comes to data skills, just less than half consider themselves to be at an intermediate level for data analysis and data visualisation. Including statistics, these are the areas in which less than 1 in 10 data journalists consider themselves to have no skills. Data wrangling, scraping, and machine learning is where over half of data journalists consider themselves to have no skills at all or view themselves as novices. For the latter, only 22% felt they are at an intermediate or advanced level.

How would you describe your skill level in each of the following?



Number of respondents differ for each entry. Hover to see the total respondents for each.

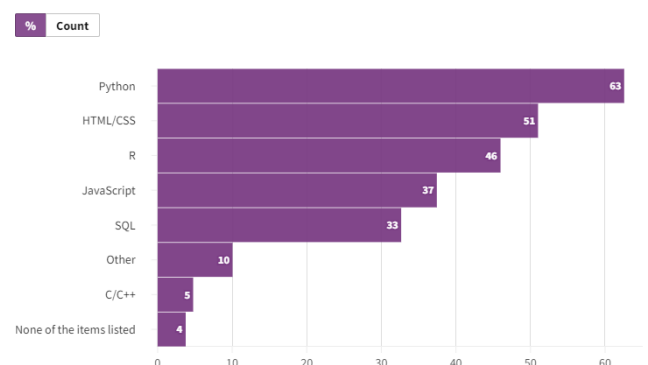
● A Flourish chart

## Programming

As we see in the next section on work practices, programming is not a mainstream task for data journalists (chosen by 29% of respondents). We asked a couple of questions about programming practices to respondents who indicated that they used coding or develop data-driven applications for their work.

The most used programming language is Python (63%), followed by HTML/CSS (51%), and R (46%). Both R and Python are extremely popular for statistical analysis, with the former having been created with that purpose in mind, and the latter being a general-purpose programming language. Python being more popular than R is also a reflection of the general community of users of each language, where Python is more widely known than R. Though JavaScript and SQL are less common, they are used by 37% and 33% of data journalists who use programming languages for work.

Which programming languages do you use for your work, if any? (Select all that apply)



This question was shown to only those who answered they code/programme for work, or develop data-driven application (n = 402). Of those, n = 398 took the question. Because this question was a multiple choice one, the percentages add to over 100. The total number of entries is n = 988. The mean number of entries per respondents is 2.48, the median 2, and the standard deviation 1.20.

● A Flourish data visualization



In terms of frequency, the vast majority uses programming daily (59%). Experience in programming varies from having more than 16 years (18%) to having between 3-5 years (28%), or 1-2 years (18%). This reveals within data journalism there is a mixture of veterans and individuals who have been programming for relatively much less time. Once again, a self-taught path is the biggest winner in terms of learning mode (81% choosing one or both), followed by higher education (36%). Workplace training (27%) and bootcamps (27%) are the least popular places for learning how to code. data journalists who use programming languages for work.

## Graphic tools

We built on a Google News Lab report question exploring data visualisation tools in news organisations. Most data journalists work in companies where they are mostly reliant on external software to produce visualisations (33%). Our findings show that in-house software is the reality for only around 11% of data journalists, which is in stark contrast with the 2017 findings from the Google News Lab report, although they only focused on mapping tools.

Whether the difference in choice relies on affordability issues for in-house tools, or the development of better, more customisable, and more affordable external software remains unknown. Even very large companies of 500+ people do not opt for in-house tools much more than medium-sized companies of 50 to 99 individuals, suggesting that price might not be playing a role in determining the nature of the chosen visualisation tool.

## Tools

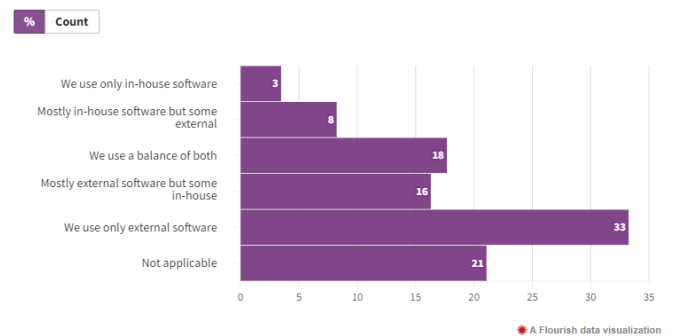
If you are a data journalist, chances are that you will have had to use Excel and/or Google Sheets -- the two most well-known spreadsheet software in existence. Largely uncontested in terms of tools for data storage, manipulation, and analysis, the former is used by 3 in 4 data journalists and the latter by 3 in 5.

The third, fourth, and fifth most popular tools are software that facilitates data visualisation: Datawrapper (37%), Flourish (32%), and Tableau (27%). Programming languages Python (26%) and R (20%) appear again, followed by OpenRefine (15%).

Undoubtedly the list of tools used by data journalists for data analysis and data visualisations is much longer (for example mapping software or network analysis software). Yet, the picture portrayed here is of a pretty undiversified landscape in terms of tools most used globally.

### Which of the following applies at your news company/organisation in terms of graphic tools, if any?

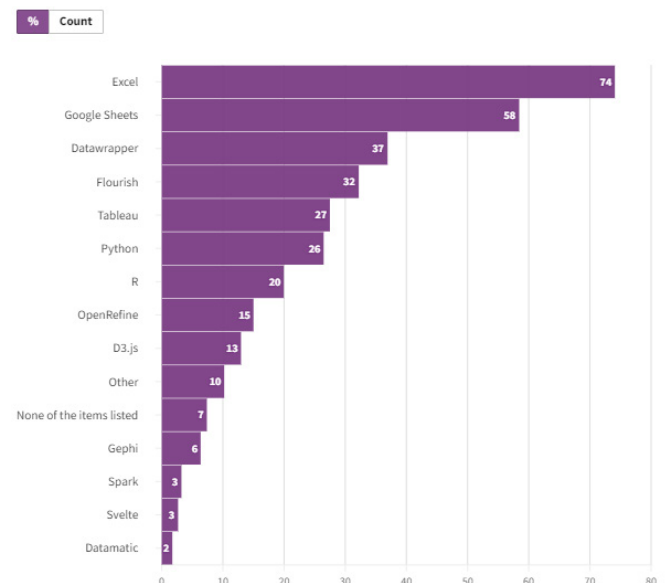
Number of respondents: 1239



• A Flourish data visualization

### Which of the following data analysis or visualisation tools do you use for your work, if any?

Number of respondents: 1273



Because this question was multiple choice, the percentages add to over 100.  
The mean number of tools used is 3.35, the median 3, and the standard deviation 2.

• A Flourish chart



## Training / upskilling / demanded training

In the survey, our aim was to find out what topics data journalists received training in, what they would like to upskill on, and what areas educators and trainers were asked to deliver training in.

Data visualisation and data analysis are the two areas where more than half of data journalists would like to upskill despite just less than half receiving training. While machine learning is what data journalists have been trained the least in (15%), it is also the topic most data journalists want to upskill in (52%). There appears to be a disconnect between the areas in which data journalists desire to be trained in, and the ones in which they actually receive training in, with the technical areas of statistics, wrangling, and scraping being the most prominent. The areas in which educators see demand for training happen to be inbetween the other two.

### (Select all that apply)

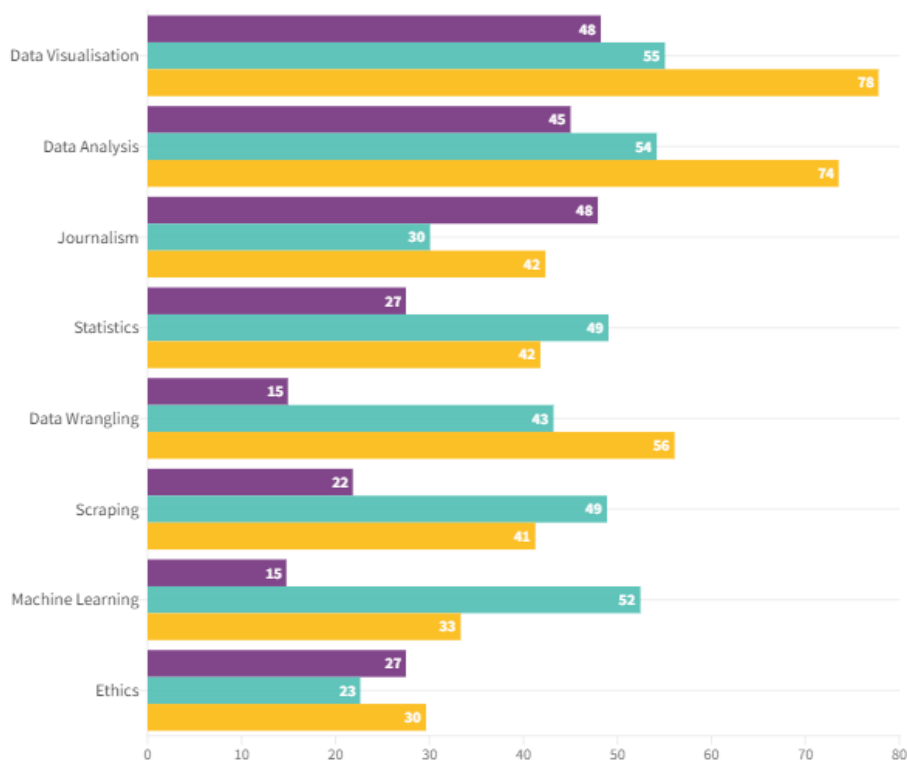
**Training** - In the last year, which of the following data skills have you been trained on, if any?

**Upskilling** - Which of the following would you like to receive upskill training on, if any?

**Demanded training** - In which areas are data journalists / data journalism students predominantly demanding training?

**All** Training Upskilling Demanded Training

■ Training ■ Upskilling ■ Demanded Training





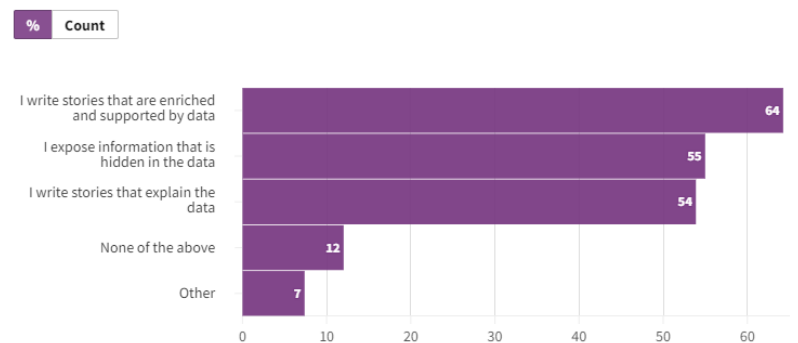
# Work Practices

## 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. Among those, we found that the most practised type of data journalism uses data to enrich the story (64%). As the authors of the 2017 Google News Lab report describe it, this type of data journalism is made of stories that do not rely on the data itself to exist but use data to support and strengthen their argument. Nonetheless, over half of respondents still participate in both investigative data journalism (e.g. Pandora Papers by ICIJ) and publishing data explainer articles.

The chart above hints that on average each person has selected two options from the answers list. Our aim was to determine how people combine the different forms of data journalism story types. How common is it to work on all three types of data stories or specialise in one type? Below we break down the counts for each possible combination. By far the most common was the selection of all three choices (37%), showing that many data journalists produce stories that use data in a variety of ways.

Which of the following do you think best applies to your work as a data journalist? (Select all that apply)

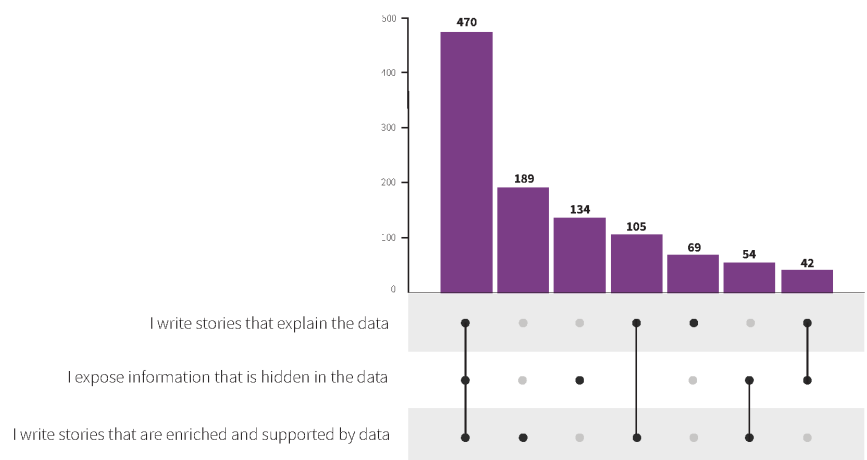


As this question was a multiple choice one, the number of entries exceeds the number of respondents. The total number of entries is n = 2451 and the respondents n = 1273

A Flourish chart

Which of the following do you think best applies to your work as a data journalist? (Select all that apply)

Number of respondents: 1273



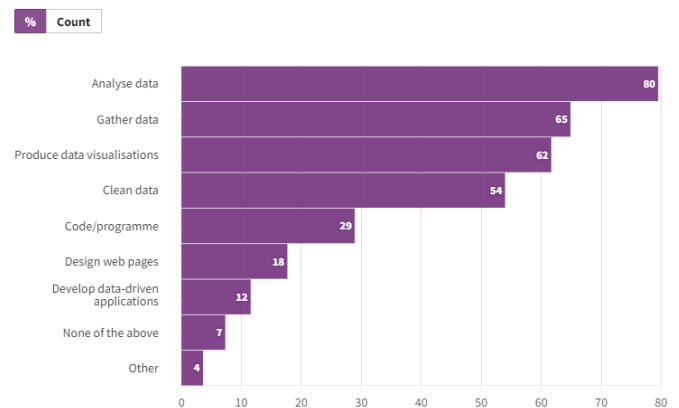




## Work tasks

We asked data journalists what data-related tasks they carry out alongside traditional reporting activities. Most data journalists analyse data in their work (80%), but data gathering (65%), data visualisation (62%), and data cleaning (54%) are relatively common. Programming for work is performed by over a quarter (29%) of data journalists. Designing web pages (18%) and developing data-driven applications (12%) remain niche tasks.

Which of the following do you carry out in your work, if any? (Select all that apply)

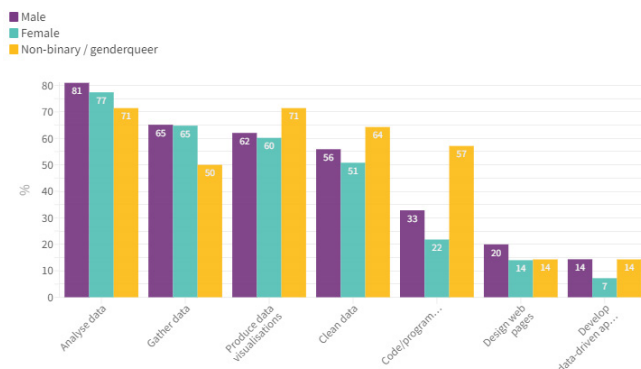


Because this question was a multiple choice one, the percentages add to over 100. The total number of entries is n = 4189, and the respondents n = 1273.

The mean number of entries per respondents is 3.29, the median 3, and the standard deviation 1.75.

• A Flourish data visualization

Which of the following do you carry out in your work, if any? (Select all that apply)



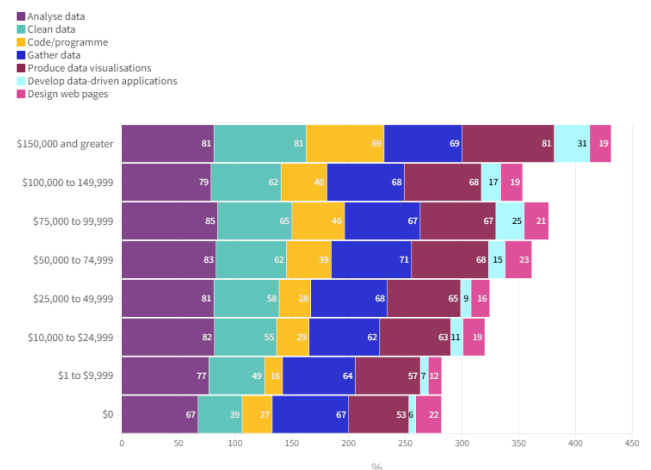
Because this question was a multiple choice one, the percentages add to over 100. The total number of entries is n = 4189, and the respondents n = 1273.

• A Flourish data visualization

As for gender, we see less women involved in programming (22%) compared to their male counterparts (33%). However, 57% of all non-binary/genderqueer respondents indicated they code/programme for work.

Another interesting data point the survey captured focused on examining the salaries of data journalists. We saw a positive correlation between carrying out more tasks and earning more money, as our last chart demonstrates. The best-paid jobs involve highly technical skills like programming and developing data-driven applications. Simply put, the more technical skills data journalists have and use in their day-to-day work, the higher the salary.

Which of the following do you carry out in your work, if any? (Select all that apply)



Because this question was a multiple choice one, the percentages add to over 100. The total number of entries is n = 4189, and the respondents n = 1273.

• A Flourish data visualization



## Scope

What scope of media did our survey results represent? The most common coverage scope is national news outlets (51%), followed by international news outlets (42%). Considering our definition of local news was quite broad, including both city, municipality, regional, and state level, it is quite remarkable to see such a small level of engagement in local news (30%). This could be an indication that data journalism may be less used in small, local newsrooms, or that the production of data-driven local news may be limited by the availability of reliable local data.

## Beat

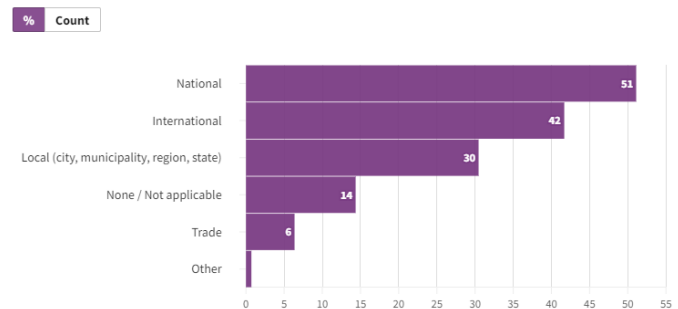
Similar to Heravi and Lorenz's survey<sup>2</sup>, or the more recent Muck Rack 2021 State of Journalism survey<sup>3</sup>, we found that Politics tops the type of beat data journalists cover (52%). This is followed by the Environment (44%). While journalists may have taken a stronger interest in environmental issues due to record-breaking temperatures and climate change-driven disasters in recent years, it is interesting to note it came ahead of the Economy (42%) and Business (33%), both traditionally data-heavy beats.

Is data journalism domain-specific? The distribution is long-tailed: the most common choice is to work on one specific beat, but many still work across several, with some selecting near to all choices.

## Medium

Altogether, 73% of respondents indicated they work for an online platform, making digital outlets the most popular medium types for data journalists. One in three say their data journalism stories are shared on social media. Meanwhile, a quarter of respondents say their data journalism stories are published in print newspapers.

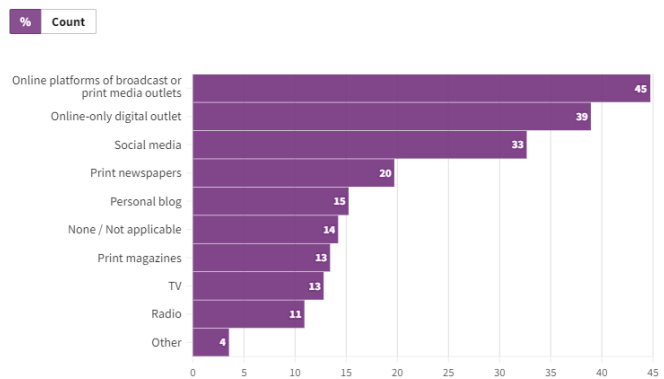
### How would you describe the scope of your coverage? (Select all that apply)



Because this question was multiple choice, the percentages add to over 100. The number of respondents is n = 1274 and the total entries n = 1843.

A Flourish chart

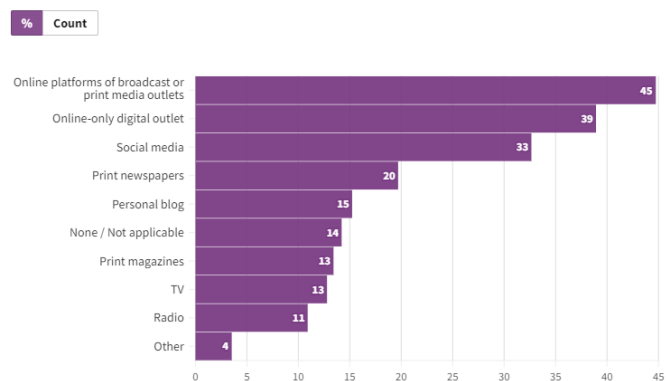
### What publication medium do you produce for? (Select all that apply)



Because this question was multiple choice, the percentages add to over 100. The number of respondents is n = 1274 and the total entries n = 2626.

A Flourish data visualization

### What publication medium do you produce for? (Select all that apply)



Because this question was multiple choice, the percentages add to over 100. The number of respondents is n = 1274 and the total entries n = 2626.

A Flourish data visualization



## Data used

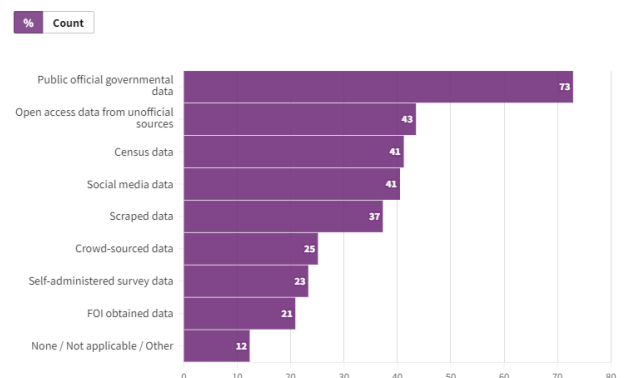
By far the most used type of data by data journalists is public official governmental data (73%). About a fifth of the time, FOI data is used (21%). On average, data journalists employed in news companies or organisations work with more types of data compared to other occupation types. Where they stand out the most, respective to the other categories, is scraped data. Instead, data that is easily accessible, like government census data, public governmental data or open-access data seems to be the preferred option for students. The category most noticeably missing by this learning group is FOI data, which instead peaks with the most professional roles, from full-time freelancing to editors and employees.

At an individual level, data journalists who earn more showed being positively correlated with working with more data types, particularly FOI, survey, scraped, and government census data. Instead, we found that company size does not exhibit any particular pattern in terms of the number of selections and preferences, with the only noticeable element being that very small companies (1 to 9 individuals), tend to have worked with fewer data types in the past year. Lastly, do we see big differences across countries? We looked at the top 20 countries by survey respondents and found that the number of data types used varies substantially. For example, Sweden and Turkey both had 15 respondents. However, our survey revealed Swedish data journalists work with a large number of data types — almost double to their Turkish counterparts.

One data type that varies substantially across countries is FOI data, which may be a reflection of differing FOI legislation from country to country. Russian data journalists ranked at the top of the survey citing usage of scraped data, social media, and public governmental data, putting them in second place for the number of data types used.

Finally, we kept an eye out for government census data in the US, as August 12th, 2021 saw the release of redistricting data from the 2020 US Census. We saw that 66% worked with census data. While this figure is impressive, and is over the average globally, it is not particularly higher than that of Canada, which will release its 2021 Census data in early 2022.

Which type of data have you worked with in the past year, if any? (Select all the apply)

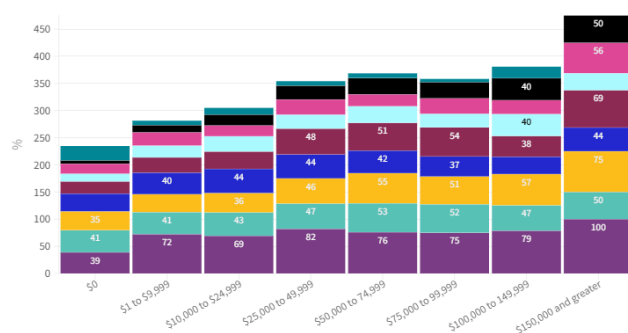


Because this question was multiple choice, the percentages add to over 100. The number of respondents is n = 1274 and the total entries n = 4039.

A Flourish data visualization

Which type of data have you worked with in the past year, if any? (Select all the apply)

Public official governmental data Open access data from unofficial sources Census data  
Social media data Scraped data Crowd-sourced data Self-administered survey data  
FOI obtained data None / Not applicable / Other

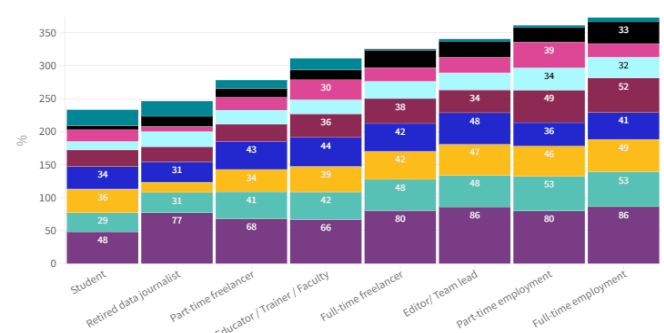


Because this question was a multiple choice one, the percentages add to over 100.

A Flourish data visualization

Which type of data have you worked with in the past year, if any? (Select all the apply)

Public official governmental data Open access data from unofficial sources Census data  
Social media data Scraped data Crowd-sourced data Self-administered survey data  
FOI obtained data None / Not applicable / Other



Because this question was a multiple choice one, the percentages add to over 100.

A Flourish data visualization



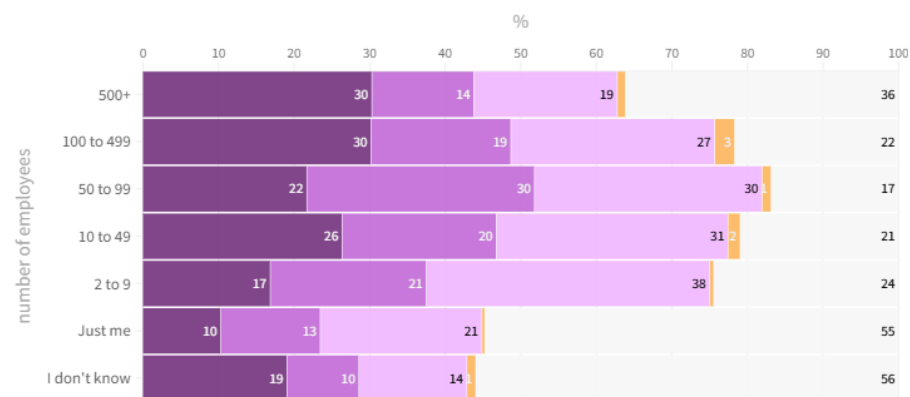
## Dedicated data unit

We found that 22% of data journalists work in a dedicated data unit, with more working in organisations that only seldom publish data-driven projects (25%). Our figure is much smaller than that of Heravi and Lorenz's paper<sup>2</sup>, which indicated just less than half of their respondents worked in dedicated data units. We see our result as a sign of the lack of resources to employ dedicated data journalists or to formally set up a data unit within the company. We found that the larger the company, the higher the share of dedicated data units, peaking at 30% in companies with over 500 employees.

### Which of the following best describes the way data journalism is performed at your news organisation, if any?

Number of respondents: 1243

- Dedicated data unit/blog/section/desk/team
- No dedicated data unit, but regularly publish data driven projects
- Sometimes publishing data driven projects
- Other
- Not applicable

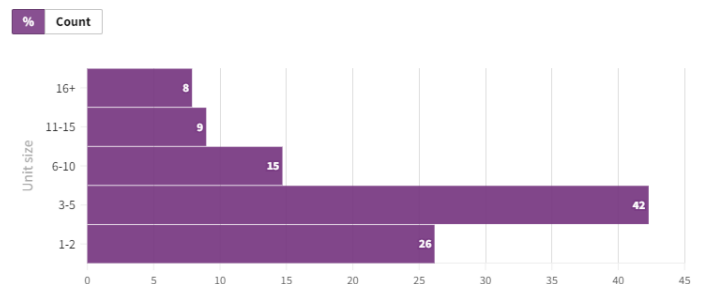


A Flourish data visualization

Within the dedicated data units, the vast majority is comprised of 3 to 5 individuals, followed by 1 to 2. This indicates that within most companies, you are likely to find one or few specialised data journalists, with the share slowly decreasing as we step up the data unit size. Among the largest dedicated data units, we found journalists from The Economist, The New York Times, Reuters, The Financial Times, and the BBC, a reminder that such big data units are possible in mostly well-established media organisations.

### How many people work in your dedicated data unit/blog/section/desk/team?

Number of respondents: 279



This question was shown to those who answered "Dedicated data unit" in the previous question. Out of 281 eligible respondents, 279 took the question.

A Flourish chart

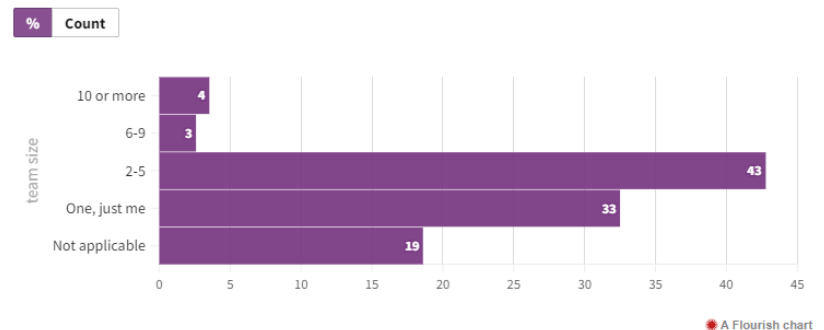


## Project team size

Some data journalism projects require large teams, sometimes even across newsrooms, to see the light of day (e.g. The Pandora Papers, the largest data investigation to date). Others, instead, are the output of a single data journalist (e.g. Following the Science, by Jeff MacInnes for The Pudding). We found that the most common project team size is small. Data journalism stories are mostly produced by small teams of 2 to 5 individuals (43%). Larger teams are quite rare (7%), while a third of data journalists usually work alone (33%).

### How many people work on the same project as you, on average?

Number of respondents: 1274

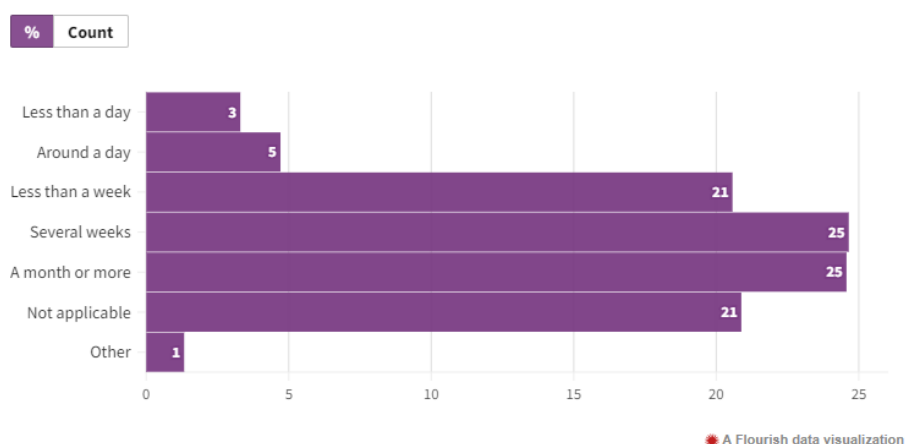


## Project duration

How long does it take to complete a data journalism project? Our survey reveals around half of data journalists said they required several weeks to over a month to complete their last project (50%). Meanwhile, a fifth (21%) completed their projects in less than a week. This is a very different finding from Google News Labs' 2017 report<sup>1</sup>, where they saw 49% delivering a project in a day or less (compared to our 8%). Although Google News Lab's survey sample focused specifically on the US, the UK, Germany, and France, our survey showed different results, even when examining our figures by country. A difference in sampling strategies might account for the observed variations, where the Google News Labs report focused on the views of data journalists and editors working within news organisations. Nonetheless, even by filtering out freelancers, students, and educators, we still do not see many variations in the results. We interpret this as a potential indicator that time pressures for data journalists in newsrooms might have reduced over the years, although other explanations are possible. As the field matures, the desire to produce more impressive designs and more sophisticated data stories may also be a motivating factor for data journalists.

### What was the duration of your last project, from pitch to completion?

Number of respondents: 1274





## Collaborations

In the span of a year, we found that 27% of data journalists have taken part in a collaborative project. We used the framework proposed by the Centre for Cooperative Media<sup>4</sup> to study these collaborations, organising them in six categories, defining the duration and production type of the project. Around half of respondents indicated participating in collaborations where content production was separate. This type of collaboration is described as fruitful for small news organisations trying to expand their reach or recognition. A similar figure was found for collaborations where the content was co-created. This is a resource-efficient approach and suits investigative reporting as well as other newsrooms with similar audiences. The least common selection was of projects where content, data, and resources were shared at the organisational level, such as in the case of the ICIJ for The Pandora Papers.

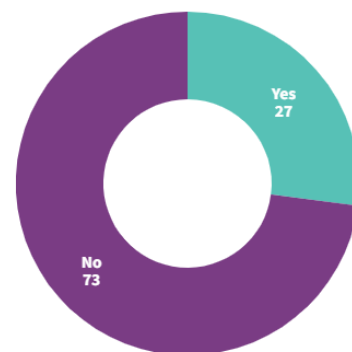
Regarding the collaboration duration, in all instances, the majority were one-time projects (61% altogether), although ongoing deals are not all that uncommon (39%).

The most widely perceived benefit of the collaboration was the opportunity to open up to wider audiences (57%), and the possibility to do better work than alone (53%). Data journalists also indicated the positive effect of learning from one another (46%) and filling existing skill and expertise holes within their newsrooms (44%).

On the other hand, different levels of tech expertise were considered the greatest challenge to the collaboration (44%), followed by differing newsroom cultures (42%). The cost of coordination and training, as well as the need to balance power dynamics, were not as commonly seen as challenges, although still reported by one in five data journalists.

### In the last year, have you collaborated with other news companies/organisations on a data journalism project?

Number of respondents: 1253

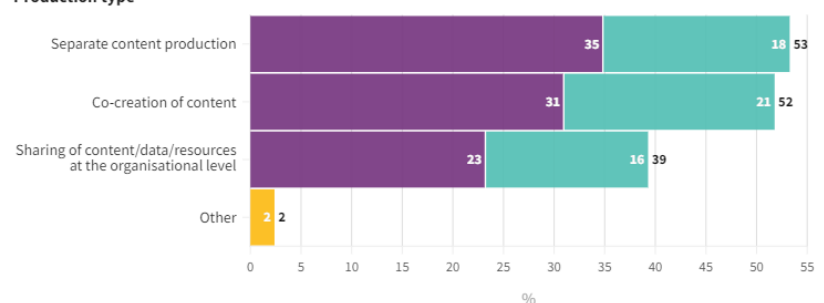


A Flourish data visualization

### If yes, which of the following best describes your collaboration, if any? (Select all that apply)

**Duration**  
■ One-Time  
■ Ongoing  
■ Other

#### Production type



This question was only displayed to those who answered "Yes" in the previous question. Of 339 eligible respondents, n = 336 took the question. The number of entries is n = 493. Because this was a multiple choice question, the percentages add to over 100.

A Flourish data visualization



# Challenges

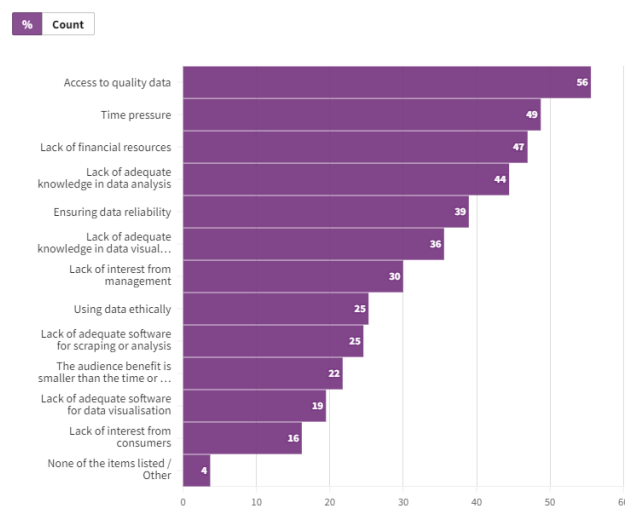
## Challenges to producing more

Building on Heravi and Lorenz's data journalism survey, as well as the Google News Lab report, we investigated the main hurdles in generating more data journalism. Our extensive answer selection was a combination of the options offered by the two reports. This was done to give a comprehensive list of possibilities to our respondents, and our findings reflect, where applicable, the results of the two studies.

The survey findings showed that the top three barriers for data journalists were access to quality data (56%), followed by time pressure (49%) and lack of financial resources (47%). In fourth place we found a lack of adequate knowledge in data analysis (44%), suggesting that data journalists could do a better job with more training. Similarly, the fifth and sixth most common hurdles were ensuring data reliability (39%), followed by a lack of data visualisation knowledge (36%).

Generally, a lack of knowledge in a specific area is seen as a bigger struggle than a lack of adequate software. On another note, not many data journalists see a lack of interest from consumers, or a poor return on investment in terms of audience benefit as a challenge, showcasing data journalists' belief in the value of their work.

**Which of the following do you consider real challenges to producing more data journalism? (Select all that apply)**



Because this question was multiple choice, the added percentages go over 100.  
Number of respondents is n = 1274.





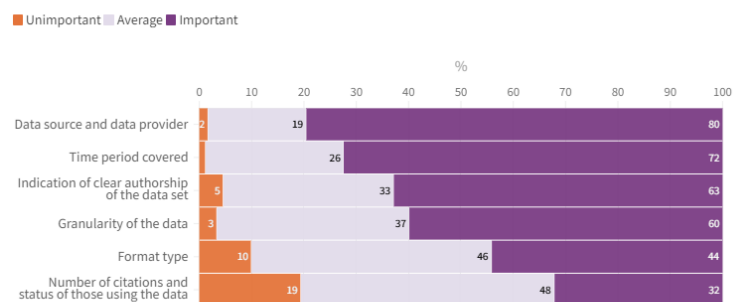
## Data access and quality

How data journalists rate the quality and access to local and national data varies substantially around the world. The table below offers a breakdown by variable type for all the countries with a minimum of 10 respondents, countries ordered by the number of respondents. Generally, local data is overall poorer both in terms of quality and access. The Scandinavian countries stand out by ranking the highest for local and national data. However, countries in other regions, such as Asia and the Middle East, paint a less promising picture for data quality and access (e.g. Egypt, Turkey, and Pakistan).

## Dataset properties

Having a clear source and provider was considered the most important feature (80%) in determining the usefulness of a dataset. Meanwhile, the number of citations and status of those using the data, or in other words, how popular a dataset is and who else has used it, was the least useful criteria (32% rated it as important). The findings signal that data journalists want to use objective parameters to determine a dataset's quality rather than external cues.

How do you consider each of the following, when determining the usefulness of a dataset?



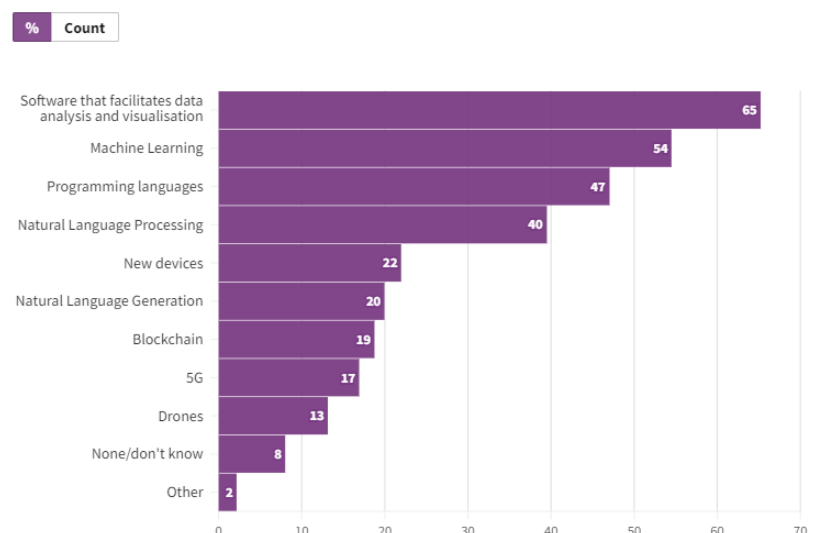
Number of respondents differ for each characteristic. Hover over to see the number of respondents for each.

A Flourish chart

## Tech opportunities

What do data journalists see as the most promising technologies for data journalism? The top choice is software that facilitates data analysis and data visualisation (65%), which shows how central these tasks are to data journalism. Over half of data journalists selected machine learning, followed by programming languages (47%) and Natural Language Processing (40%).

In your opinion, which of the following technologies is likely to become most important for data journalists to harness over the next five years? (Select all that apply)



Because this question was multiple choice, the added percentages go over 100. Number of respondents is n = 1248.

A Flourish chart

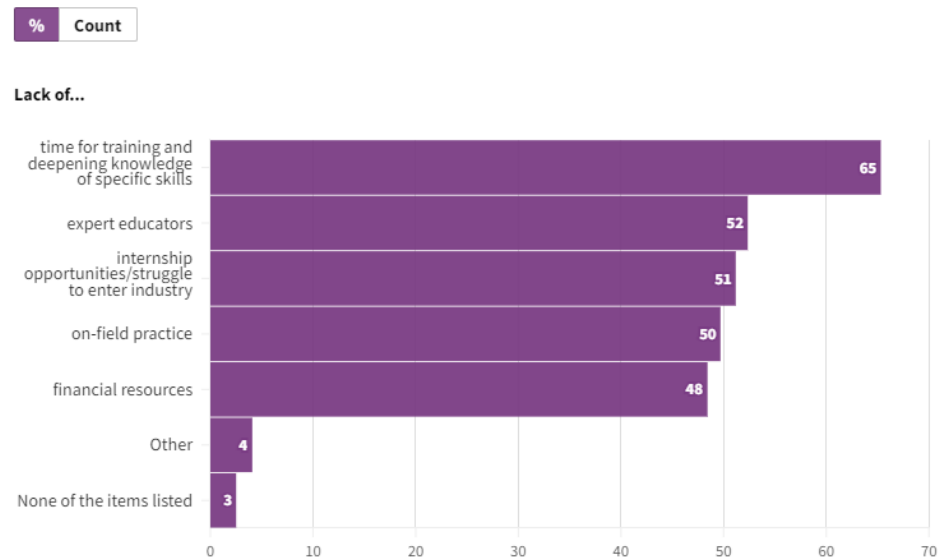




## Barriers to learning

We asked educators specifically to tell us what they viewed as the greatest hurdles for learning data journalism. Lack of time was the most common selection (65%), followed by a lack of expert educators (52%) and the struggle to enter the industry (51%). Altogether, nearly half of educators saw a barrier in each of the listed options, revealing how many factors hinder learning data journalism.

### What do you believe are the greatest barriers to learning for data journalists/data journalism students? (Select all that apply)



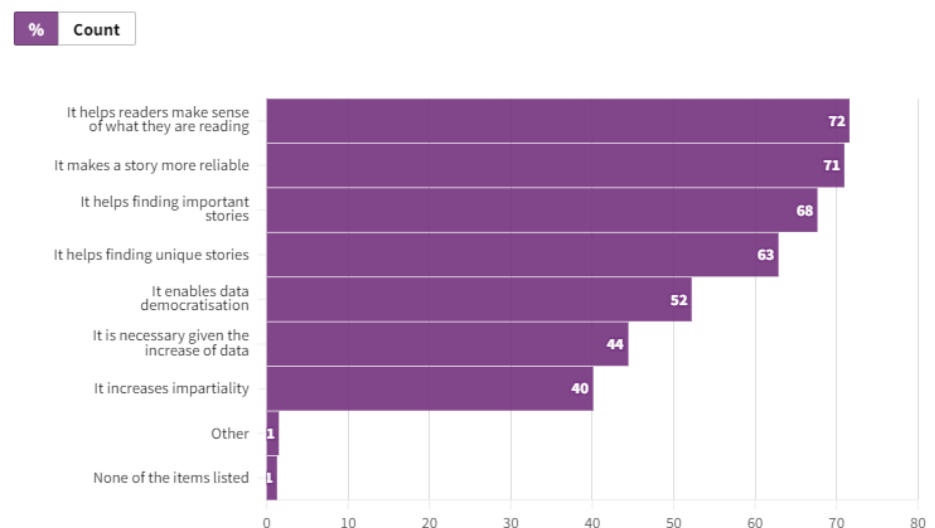
Because this question was multiple choice, the added percentages go over 100.  
Number of respondents is n = 1274.

A Flourish chart

## Value of data journalism

What does data journalism bring to the field at large and to the audiences that consume it? About 7 out of 10 data journalists who responded to the survey said data journalism offers reliability and contextualisation of a story. However, data journalism also facilitates finding (68%) relevant or unique (63%) stories. Less than half indicated the need for data journalism to operate due to much information existing in data format (44%), and even less saw data journalism as a way to increase impartiality (40%).

### Which of the following would you say applies to data journalism, in terms of its value and contribution? (Select all that apply)



Because this question was multiple choice, the added percentages go over 100.  
Number of respondents is n = 1274.

A Flourish chart



# Pandemic and work

The COVID-19 pandemic has been a central topic for data journalists across the globe, challenging them to provide coverage for it, navigating hard to explain mechanisms behind contagion, variants, and protection, not to mention feeling the consequences of it at work and in their private lives.

ICFJ and Tow Center for Digital Journalism put together a study and in-depth research examining the impact of the COVID-19 pandemic on journalism. Their first report showcases results from a spring 2020 survey spanning over 125 countries. The research reflects a grim picture of job cuts, the struggle to battle disinformation, and a decline in mental health. On the positive side, however, they found that journalists believed they saw an increase in trust in journalism.

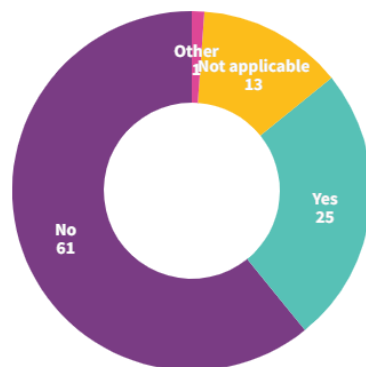
In this survey edition, we dedicated one section to the pandemic and asked data journalists to tell us how they have been affected by it.

## Involvement in data journalism

We found that one quarter of our respondents actually became involved in data journalism as a result of the pandemic. On the one hand, this is the result of newsrooms allocating manpower to the coverage of the pandemic, which inherently data-driven. On the other hand, it could be that more journalists have become interested in data journalism due to the rise in the popularity of charts and data visualisations. Alternatively, perhaps they saw a need to participate in the coverage using techniques specific to data journalism.

### Have you become involved in data journalism as a result of the pandemic?

Number of respondents: 1262



A Flourish chart



## Coverage of the pandemic

Over the course of 2021, more than half of data journalists surveyed (56%) have worked on a project that covered the COVID-19 pandemic in some fashion. Although news coverage and angles can differ substantially, this figure suggests the current pandemic is the most commonly shared topic covered by data journalists worldwide.

## Impact of the pandemic

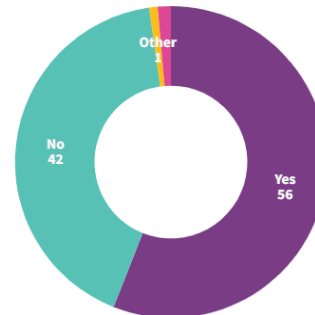
When examining the impact of the pandemic on data journalism, the most widespread views are that it has strengthened data journalism as a field (46%) and increased audience data literacy (43%). Well over one-quarter of data journalists also believe they've seen an improvement in access to data (28%).

Another positive effect, shared by 9% of respondents, is that the pandemic has made the job of data journalists easier. On the other end of the spectrum, however, 19% believe that the pandemic has made the jobs of data journalists more difficult, and 9% think it has weakened the field in some measure. We recognise that the phrasing of the options leaves things up for interpretation. For instance, one could deduce that the pandemic has made the job of data journalists both easier and harder in different ways. For example, it could have improved the working conditions of employees, but challenged their mental health or increased time pressure to deliver, or vice versa. By examining the results at face value, we conclude that data journalists view COVID-19 as having had a predominantly positive impact on establishing the field and informing audiences.

### Have you worked on a covid related project in 2021?

Number of respondents: 1250

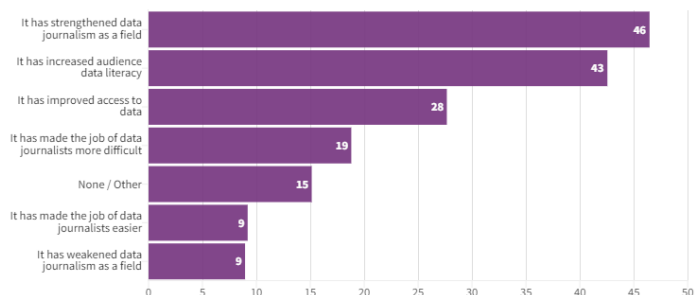
% ■ Yes ■ No ■ Other ■ I don't know



• A Flourish chart

### In what way would you say the pandemic has impacted the field of data journalism in your region, if any? (Select all that apply)

% Count

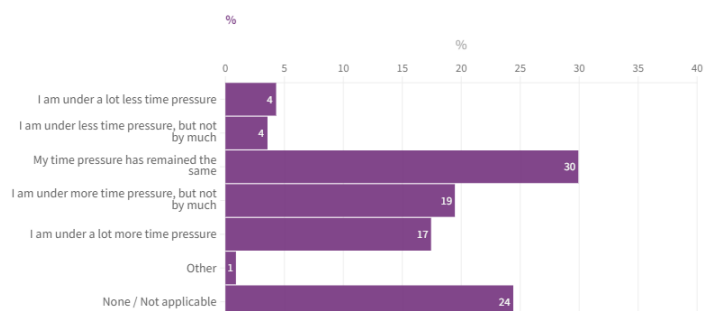


Because this question was multiple choice, the added percentages go above 100. The total number of entries is n = 2074, and the total number of respondents n = 1229.

• A Flourish chart

### How has the pandemic affected your...

Resources Time pressure Workload Workmode





Rather than exploring their views on how the pandemic has affected data journalists at large, we here asked them to tell us about their personal work-related experiences. The good news is that when it comes to resources, workload, and time pressure, the majority deems themselves unaffected (all between 26% and 30%). The bad news is that more data journalists have otherwise been negatively rather than positively affected.

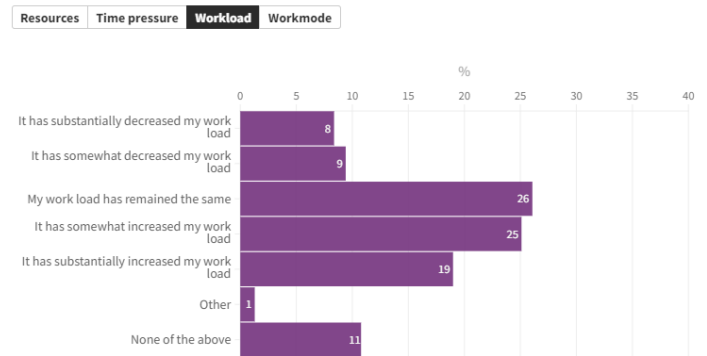
The survey showed 15% saw a big decrease in resources, and 17% a somewhat decrease in resources, totalling 32% of data journalists. But the biggest gap between positive and negative options is with time pressure: while 30% have been unaffected, 36% have seen their time pressure somewhat (19%) or highly (17%) increase. Similarly, 44% saw an increase in their workload, although 13% have instead seen a decrease.

In terms of work mode, the vast majority (40%) has switched to working from home, and 32% instead converted to a hybrid work policy. Finally, 19% continued working the same way they used to, may that have been from home (e.g. for freelancers) or from the office.

## Skill improvements due to the pandemic

Finally, we asked our respondents to tell us where they felt their skills improved as a result of the pandemic. At the top of the charts, we found data analysis at 40%, followed by journalism at 39%, and data visualisation at 36%.

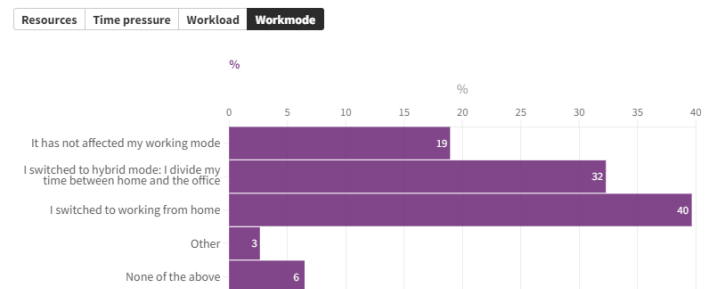
### How has the pandemic affected your...



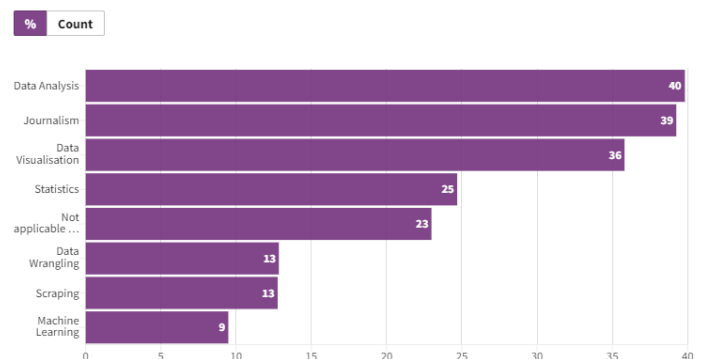
Hover over to see number of respondents for each section.

A Flourish chart

### How has the pandemic affected your...



### In which of the following would you say your skills have improved as a result of the pandemic, if any? (Select all that apply)



Because this question was multiple choice, the added percentages go above 100. The total number of entries is n = 2417, and the total number of respondents n = 1223.

A Flourish data visualization



# Thank you

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

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

And to our supporters, for offering tools, goodies, and insights.



# Cited Work

Heravi, Bahareh R., and Mirko Lorenz. "Data Journalism Practices Globally: Skills, Education, Opportunities, and Values." *Journalism and Media* 1, no. 1 (2020): 26-40.

MuckRack. "The State of Journalism 2021".

Posetti, J., Bell, E., & Brown, P. (2020). *Journalism & The pandemic – A global snapshot of impacts*. International Center for Journalists.

Rogers, Simon, J. Schwabish, and D. Bowers. "Data journalism in 2017." Google News Lab (2017).

Stonbely, Sarah. *Comparing models of collaborative journalism*. Center for Cooperative Media, Montclair State University, 2017.