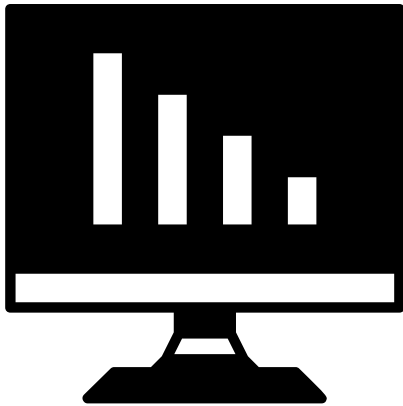# What is data-driven journalism (ddj)

Data-driven journalism is a special form of

1. **research,** which aims to discover stories in datasets
2. **interpretation** of findings, which uses statistical methods to gather insight
3. **presentation**, which illustrates key findings visually and sometimes interactively

Types of data-driven stories

NEWS STORY
- Gives context for current events using data and graphics
- **Format**: News, short articles

Example: How privilege shapes learning outcomes in India – mint India

INVESTIGATION
- Uses exclusive data and leaks to uncover new information
- **Format**: Elaborate longreads, headline news

Example: Panama Papers — Süddeutsche Zeitung & ICIJ

DATA ANALYSIS
- Discovers and explains trends, patterns and outliers in a dataset
- **Format**: News, feature or explainer

Example: Murder in The Arms of Marriage – InfoTimes Egypt

SERVICE
- Lets the audience interact with a dataset and find their own stories
- **Format**: Interactive graphics

Example: [What would a powerful earthquake feel like where you live? Search our map](#)
— Los Angeles Times

VISUAL STORYTELLING
- Centers data visualization and multimedia elements in order to tell a story
- **Format**: Longread or feature, explainer

Example: [The race to save the river Ganges](#)
— Reuters Graphics

Why use data-driven journalism?

CONTEXT
- go beyond quotes and anecdotes
- aggregate data from many individual cases
- show systemic problems in context

INDEPENDENCE
- be independent from third-party studies
- fact-check claims for yourself

PRESENTATION
- easily present large amounts of information
- enhance long, dull text articles
- let the audience find their own stories in interactive data

EXCLUSIVITY
- report on stories other journalists can't find
- provide new angles to existing discussions

Fallacy of neutrality

Data is not objective and neutral! It is collected, analyzed and communicated by humans.

Some factors that influence your data can be:

- underlying definitions
- methods used in data collection
- data selection
- data presentation
- human error
- intentional manipulation

Workflow

IDEA ① **RESEARCH** ② **CLEANING** ③ **ANALYSIS** ④ **VISUALIZATION** ⑤
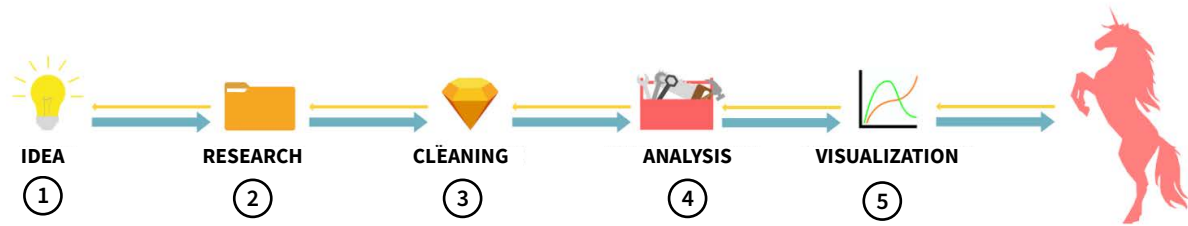
IDEA
① Formulate a hypothesis that can be investigated using data

RESEARCH
② Gather the appropriate data

CLEANING
③ Prepare your data for analysis

ANALYSIS
④ "Interview" your data to check your hypothesis

VISUALIZATION
⑤ Present your findings in an engaging and comprehensible way

---

Don't forget

Data journalism is still journalism:

You will also
- Find protagonists
- Interview experts
- Fact-Check findings

Tools

You don't have to be a programmer to be a data journalist – but it helps.

The main tools for data journalists are:

- Spreadsheets:
  - LibreOffice, Microsoft Excel, Google Sheets.
  - Tools like Workbench specifically for data analysis
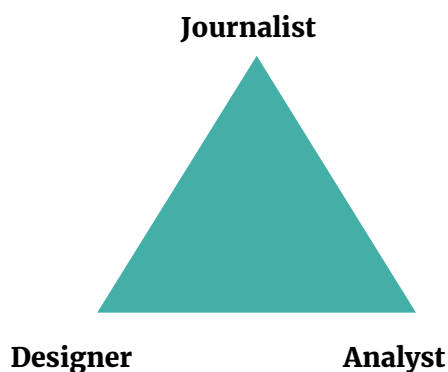- Data visualisation: Platforms like Datawrapper or Flourish
- And many more

For programmers: R, Python, and more

Skills

Journalist

Designer   **Analyst**

Data journalists need:

- **journalist**'s thinking
- **designer**'s creativity
- **analyst**'s data skills

Collaboration in the newsroom

Depending on the size of the newsroom, there might be different models for data teams. They usually fall somewhere between these two extremes:

JACK-OF-ALL-TRADES

This generalist can analyze, design and write sufficiently well to produce a data-driven story on their own
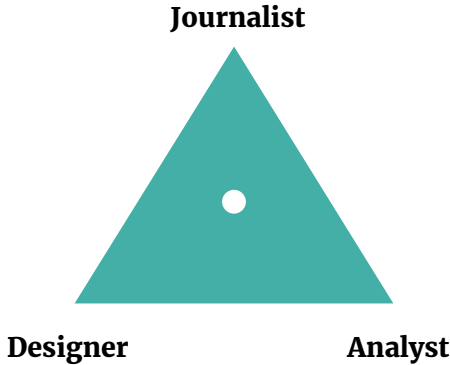
THE SPECIALISTS

In a bigger data team, each member might specialize in one aspect of data-driven work and only be responsible for their part of the workflow

Journalist

Designer   Analyst

Journalist

Designer   Analyst