Telling Stories with Data

Insights into Data Journalism

Yannik Buhl

Twitter: @YBuhl

Wiesbaden/Mannheim, May 26th, 2021

In your terms: What defines data journalism?

What defines data journalism?

- Nothing new, per se

- #ddj: *Data Driven Journalism* (data are not a simple accessory anymore)

- Quite popular for a few years now

- "Boom" thanks to advancing technological developments (easier web-scraping, etc.)

- Strong elements of IT and statistics

Overview

1. Where do data stories originate?

2. Challenges of data journalism + examples

3. Scientists and data journalists

4. Common tools in data journalism (excerpt)

5. Popular examples and what to learn from them

6. Where to find excellent data journalism

1. How to find data driven stories?

1. How to find data driven stories?

- Going from the story to the data

- Going from the data to the story
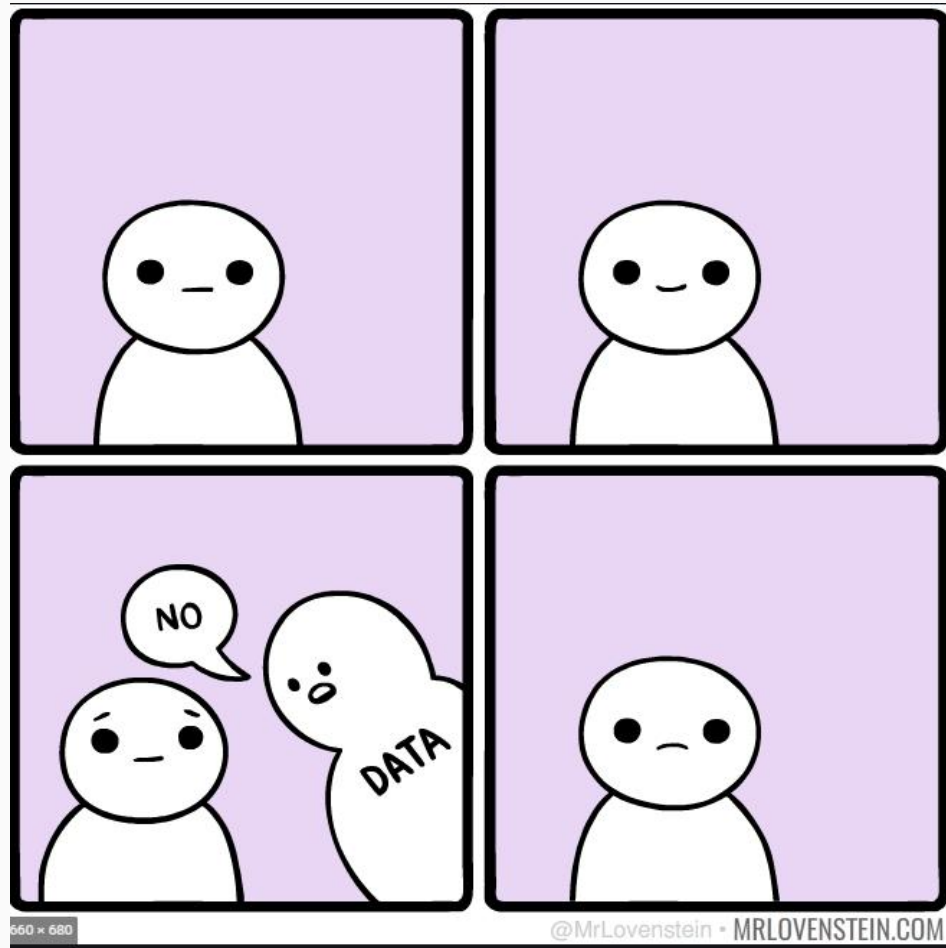
You should almost be able to read the story
without having to know that it comes from a dataset.

Caelainn Barr, Data Projects editor at The Guardian[1]

[1]In: *The Data Journalism Handbook 1*

1. How to find data driven stories?

Classic journalist work: What is a story?

- It is all about the local – keep the focus on the user!

- Relevance for society and reader most important (& explain why!)

- Need to draw readers' attention – be close to their lifes

- Fact check as much information as possible

- Think of a headline while doing your project

- Think of who your audience is & discuss with colleagues

- Example

1. How to find data driven stories?

Working a story with data

- Know the questions you want to ask the data

- Know what you should not find in there

- Count or total something, put numbers in context by proportion

- International or intranational comparisons

- Show and analyse change over time

- Transform the data to change perspective (e.g., aggregate)

- Example

2. Challenges for data driven journalism[2]

2.1 Getting the data

2.2 Data quality

2.3 Data wrangling

2.4 Data visualisation

2.5 Fundamental additional steps

---

[2] Needless to say, one often faces all challenges at once during a project.

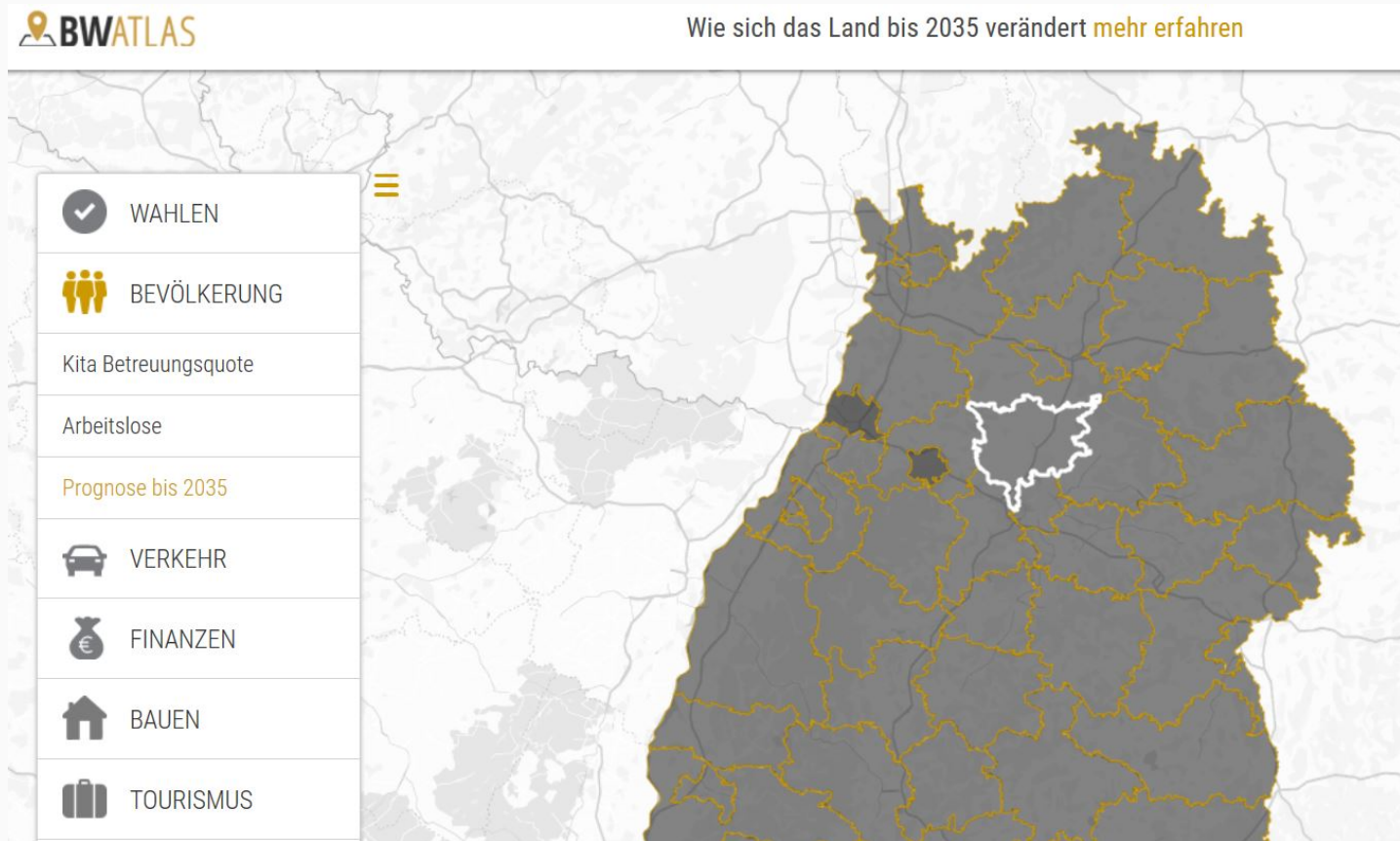# 2. Challenges for data driven journalism

2.1 Getting the data

- Where do I find data for my stories?

- Open Data (in a broad sense)

- Pay for data

- Collecting data yourself[3]

- Scrape data (on the Internet vs. PDFs/Excel)

[3] Example: Süddeutsche Zeitung / Deutschlands Mietmarkt ist kaputt

2.1 Getting the data – Example: BW Atlas

Example

2.1 Getting the data – Example: Umweltbonus

Example

Bilanz Umweltbonus

Daimler profitiert nur wegen des Smart

Von Yannik Buhl - 01. Juli 2019 - 11:00 Uhr

Der staatliche Umweltbonus soll dafür sorgen, dass endlich mehr Elektroautos unterwegs sind. Nach zweieinhalb Jahren zeigt sich: Die Regierung hat ihr Ziel weit verfehlt. Und bisher hat sich die Prämie nicht für alle Hersteller gelohnt.

Der Umweltbonus hat sich beim E-Smart bemerkbar gemacht. Foto:

2. Challenges for data driven journalism

2.2 Data quality

- What was the data generating process?

- Do they actually represent what is important for my story?

- Are there potentially erroneous data? (e.g., GPS data)

- Which alternatives do exist? Are they a better solution?

2.2 Data quality – Example: Immobilienatlas

Example

Crying about missing data
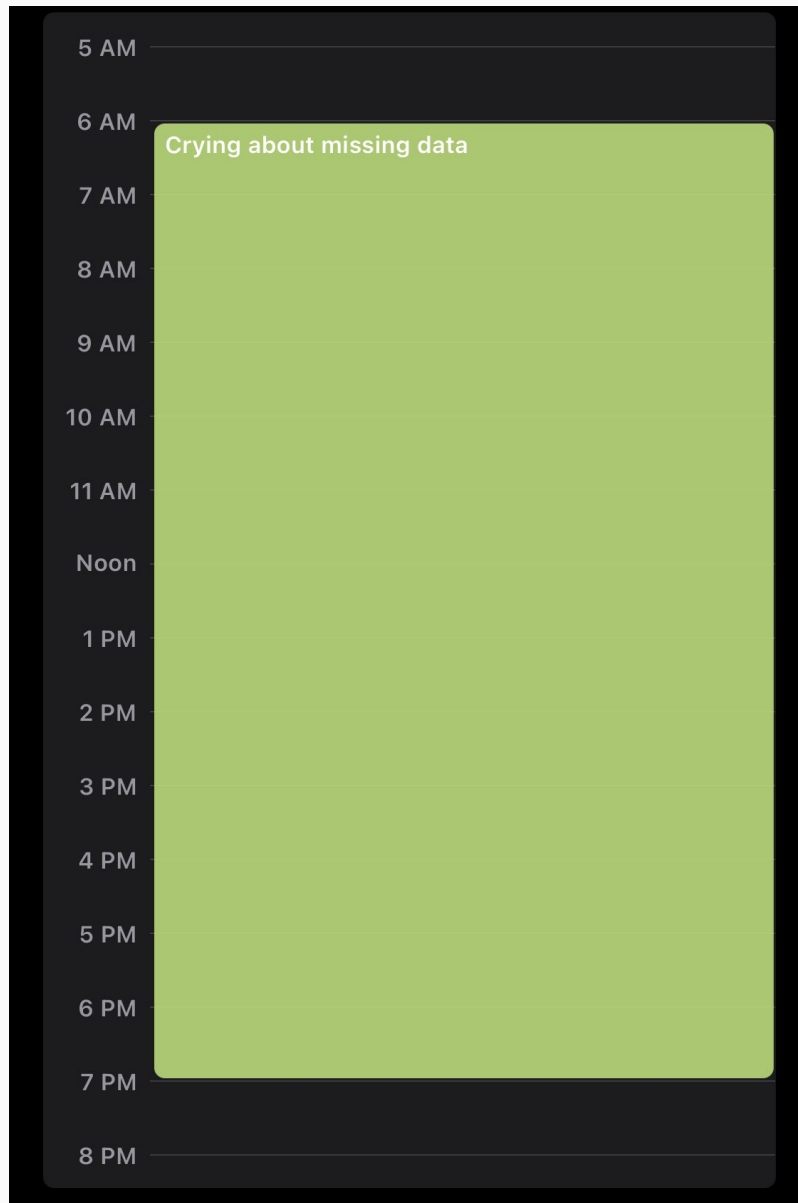
5 AM
6 AM
7 AM
8 AM
9 AM
10 AM
11 AM
Noon
1 PM
2 PM
3 PM
4 PM
5 PM
6 PM
7 PM
8 PM

2. Challenges for data driven journalism

2.3 Data wrangling

- In most cases, we deal with the worst possible data formatting

- Dealing with big data requires automatisation

- Sometimes programming skills helpful / necessary

- Calculate new variables/groups of interest

- Merging new data sets to find new stories (e.g., commuting/BW-Atlas)

2.3 Data wrangling – Example: EU Parliament Election I

27. Mai 2019, 15:31 Uhr   Europawahl

So wurde in Deutschlands Landkreisen gestimmt

Wo die CSU triumphierte, in welchen Landkreisen die Grünen stärkste Kraft wurden - und in welcher Region die AfD unter fünf Prozent blieb. Analyse und Grafiken.

Von Oliver Das Gupta und Hannes Munzinger

f   y   ◎   ✉   ▣   ✔   Artikel drucken   Feedback

Facebook   Twitter   WhatsApp   E-Mail   Flipboard   Pocket

2. Challenges for data driven journalism

2.4 Data visualisation

- How do I communicate my results comprehensively to the audience?

- UX of utmost importance (be concise!)

- Many say: biggest challenge, but for sure: the hardest task

- Highly dependent of resources in newsrooms

- Many data visualisation tools, even more ways of presenting results

- Mobile first!

- Accuracy versus understandability

2. Challenges for data driven journalism

2.4 Data visualisation

Thinking about delivering data

- To visualise or not to visualise?

- The story guides the visualisation

- People tend to stay longer when there is a data viz ...

- ... but you loose them quickly if it's too complicated

- Who is my audience – and what are their needs?

2. Challenges for data driven journalism

2.4 Data visualisation

Some thoughts about visualisation

- What does the reader need to know to understand your viz?

- Tell the world: Give access to the data & make Social Media graphs

- Make it interactive to search for your area

- Visualisations can help identify a story in the first place

- Identify and highlight the unexpected / patterns

- Discuss with others outside of your field

2.4 Data visualisation

Data journalism during the pandemic

- Importance to avoid misinterpretation

- Importance of scales, putting things into perspective

- Make a map an animation to show change over time

- Difficult: Changing data base, operationalisation, domain knowledge...

Data journalists need to continuously balance the scientific claim of being correct and the journalistic claim to be accessible. If readers need to first read a 'package leaflet' for a visualisation, it does not serve the purpose of making a complex topic understandable at first glance.

Marie-Louise Timcke, Head of Funke Interaktiv, Journocode

2.4 Data visualisation – Example
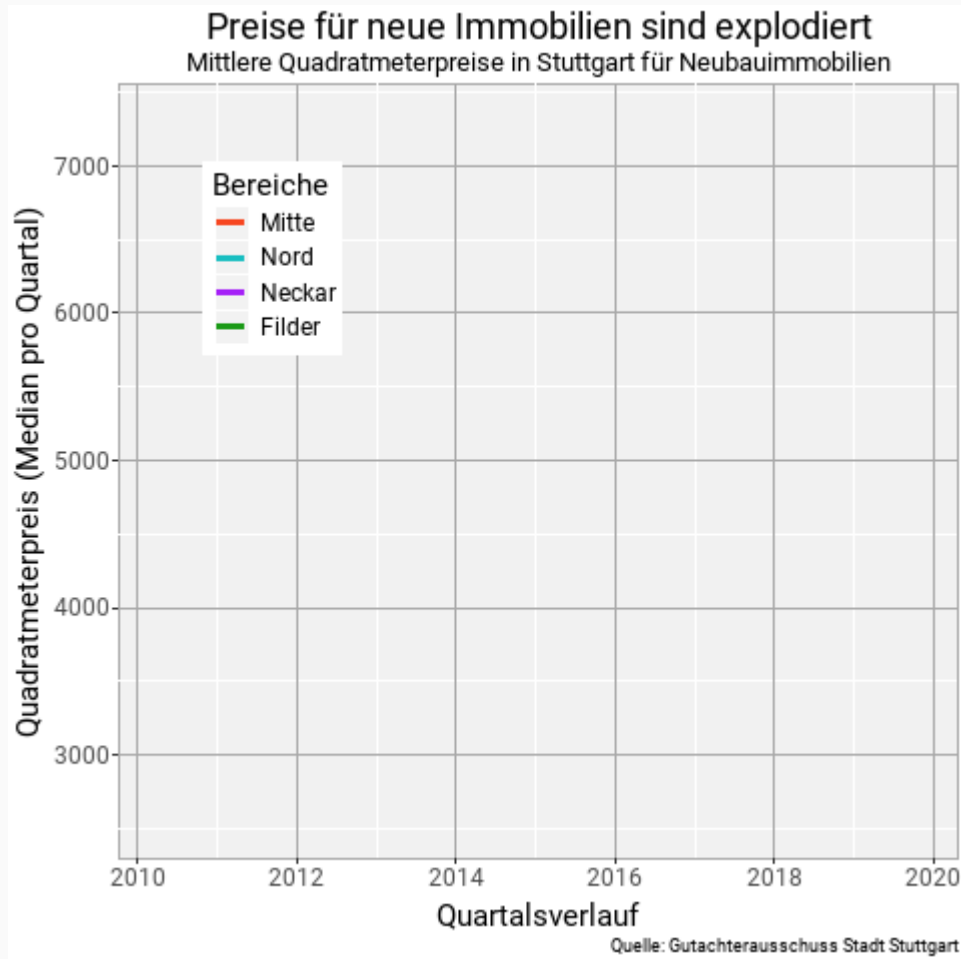
2.4 Data visualisation – Example: EU Election II

Example

27. Mai 2019, 19:30 Uhr   Europawahl

Das sind die deutschen Abgeordneten

2.4 Data visualisation – Example: EU Election III

Example

EU-Parlament

So sieht Europas Parteienlandschaft aus

181 Parteien aus 28 Ländern sind ins neu gewählte Europäische Parlament eingezogen. Dort verbünden sie sich zu verschiedenen Fraktionen. Aber wie gut passen diese Gruppierungen politisch zusammen? Eine Analyse in Grafiken.

Von Yannik Buhl, Christian Endt und Sarah Unterhitzenberger

2. Challenges for data driven journalism

2.5 Fundamental additional steps

- Embrace transparency in simple words

- *Always* explain to the reader the limitations of your analysis

- Talk to people!

    - Talk to those who are affected by the problem/topic at hand
    - Talk to experts that give a qualitative perspective on trends/effects
    - Talk to interest groups, politicians, etc. to ask for consequences

3. Scientists and data journalists

3. Scientists and data journalists

3.1 What to learn from each other?

- The importance of a compelling story (drawing readers' attention)

- High value of scientific standards

- Thinking about the boundaries of your analyses and communicate them

- Presenting and communicating results in a relevant and easy-to-understand way

3. Scientists and data journalists

3.2 Boundaries of data journalism

- Some data journalism projects reach scientific approaches

- How to explain complex data analyses (e.g., regression, causality)

- How to communicate scenarios, predictions and uncertainty (e.g., CIs)

- Need to provide lengthy methodology, but readers won't read

- **When do journalists go too far?**

4. Tools in data journalism (exzerpt)

4. Tools in data journalism (exzerpt)

- Excel (duh!?)

- Programming languages such as R & Python

- Alternatively: http://openrefine.org/ (data cleaning, wrangling, joining)

- Web-scraping tools (e.g., scraping-bot.io, scrapy.org, Tabula/PDFs)

- Data visualisation tools (e.g., Datawrapper, Flourish, Tableau)

- JavaScript (advanced web development / visualisation)

- QGIS (advanced geo data / data mapping)

- https://datajournalism.com/

- Google ...

5. Popular examples and what to learn from them

5. Popular examples and what to learn from them – Hanna & Ismail

https://www.spiegel.de/wirtschaft/soziales/diskriminierung-auslaender-werden-bei-der-wohnungssuche-benachteiligt-a-1153297.html

Studie von SPIEGEL und Bayerischem Rundfunk

Migranten werden bei der Wohnungssuche benachteiligt

Wohnungssuche mit ausländischem Namen ist schwer. Besonders stark diskriminiert werden Menschen arabischer und türkischer Herkunft. Das zeigt ein groß angelegtes Experiment in zehn deutschen Städten.

22.06.2017, 06:02 Uhr

5. Popular examples and what to learn from them – Climate change

https://www.nzz.ch/panorama/hitzewelle-schweizer-staedte-werden-immer-heisser-ld.1491209

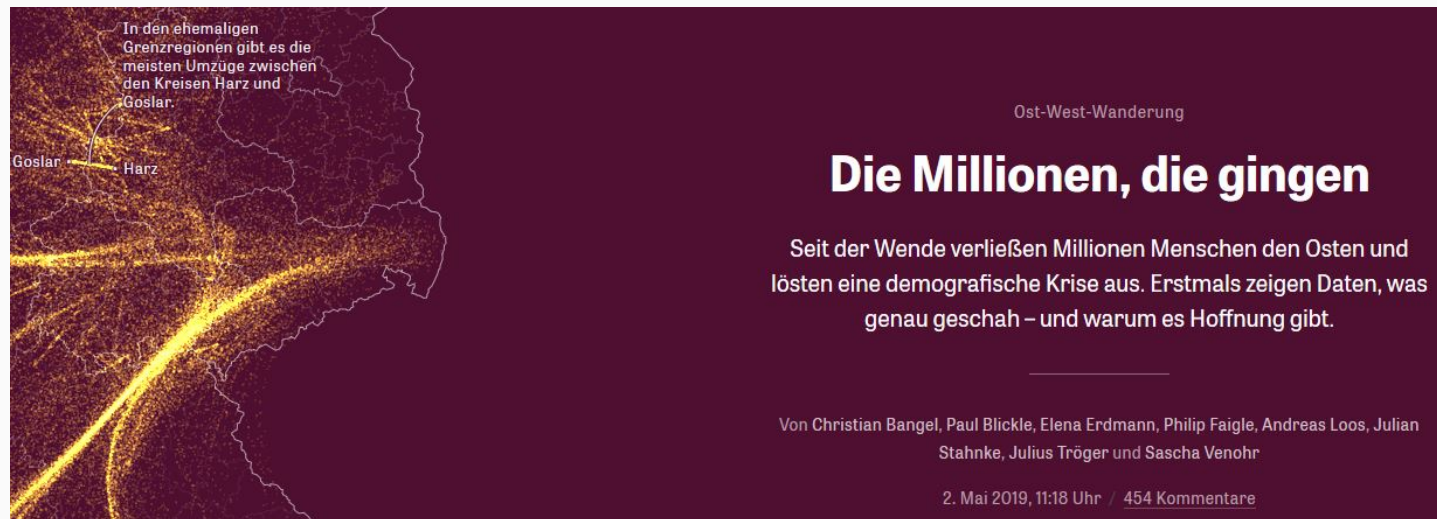# 5. Popular examples and what to learn from them – "So tickt Berlin"

https://interaktiv.morgenpost.de/berlin-an-deiner-linie/

5. Popular examples and what to learn from them – East and West Germany

https://www.zeit.de/politik/deutschland/2019-05/ost-west-wanderung-abwanderung-

ostdeutschland-umzug

6. Where to find excellent data journalism (exzerpt)

6. Where to find excellent data journalism (exzerpt)

Germany

- *ZEIT Online*

- *Spiegel*

- *Öffentlich-Rechtliche (e.g., BR)*

- *Süddeutsche Zeitung*

- *Berliner Morgenpost*

International

- *New York Times*

- *Five Thirty Eight*

- *The Economist*

- *The Guardian*

- *Neue Zürcher Zeitung*

... and lastly: Some hot topics in data journalism

- Sensor data journalism

- Automatisation

- Using machine learning

- Use satellite data