

Data Journalism with RMarkdown

```
r-ladies_global %>%
  filter(city %in% "Berlin", topic %in% "Markdown") %>%
  mutate(guest = "Marie-Louise Timcke")
```

R Markdown at R-Ladies



It's a me - Marie



data journalism student at TU Dortmund

Co-Founder of Journocode

Data Trainee at Interactive Team, Berliner Morgenpost

Rstats-Lover <3

Journocode



- Meetup, Barcamp, ddj Training
- Learning ddj techniques together
- Tutorials, Workshops
- @journocode | journocode.com

Journocode

R: Tidy Data

R | TUTORIALS

BY MARIE-LOUISE TIMCKE / ON MARCH 5, 2016 / 7 COMMENTS



Unfortunately, data comes in all shapes and sizes. Especially when analyzing data from authorities. You'll have to be able to deal with pdfs, fused table cells and frequent changes in terms and spelling.

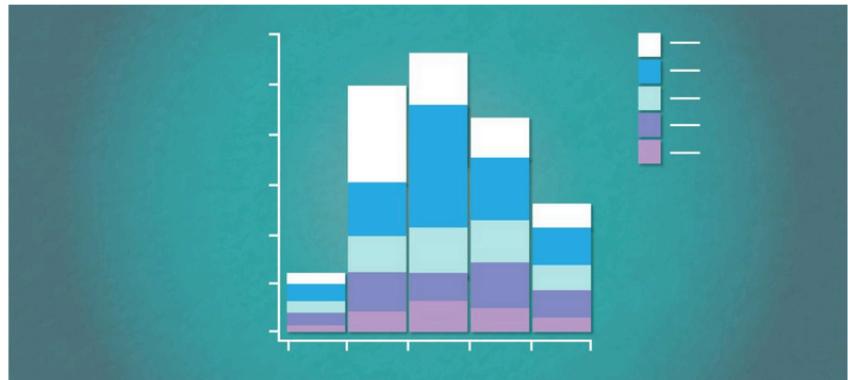
When I analyzed the [swiss arms export data](#) as an intern at [SRF Data](#), we had to work with scanned copies of data sheets that weren't machine-readable, datasets with either french, german or french *and* german countrynamnes in the same column as well as fused cells and changing spelling of the categories.

[Full post](#)

R: plotting with the ggplot2 package

R | TUTORIALS

BY MARIE-LOUISE TIMCKE / ON MARCH 2, 2016 / 5 COMMENTS



While crunching numbers, a visual analysis of your data may help you get an overview of your data or compare filtered information at a glance. Aside from the built-in *graphics* package, *R* has many additional packages to help you with that. We want to focus on [ggplot2](#) by Hadley Wickham, which is a very nice and quite popular graphics package.

Ggplot2 is based on a kind of statistical philosophy from a book I really recommend reading. In [The Grammar of Graphics](#), author Leland Wilkinson goes deep into the structure of quantitative plotting. As a product, he establishes a rulebook for building charts the right way. Hadley Wickham built ggplot2 to follow these aesthetics and principles.

[Full post](#)

Journocode

The screenshot shows the homepage of Journocode. At the top, there's a navigation bar with a squirrel icon, followed by links for HOME, DDJ TOOLS (which is highlighted in teal), DICTIONARY, and EVENTS. Below the navigation is a main content area with a dark background. It features a search bar with placeholder text "Search..." and a magnifying glass icon. A central text block reads:

A collection of tools for all your journocoding needs. Filter your search to find just the kind of tool you are looking for right now. Some of the tools are tagged with the *golden acorn*. Those are some of our favourites. Check them out, you might like them too. If you have a suggestion, tell us via [twitter](#), [facebook](#) or [mail](#).

Below this, there are several tool cards arranged in a grid:

- BitBucket**
Code sharing platform, free private and public repositories for up to five contributors
- Carto**
Create interactive, customizable maps right in your web browser
- CodePen**
Online editor and code sharing platform
- ColorHexa**
Find the right colors, create schemes and gradients
- Datamatic**
Editable data visualizations for the web
- Datawrapper**
A handy web tool to create interactive charts, maps and more
- Excel Geocoding Tool**
Get longitude and latitude data from addresses via Excel
- Forensically**
Check images for hoaxes
- Socially**
Beta Open File Help [Like](#) Share [Email](#) Magnifier [Enhancement Histogram](#)

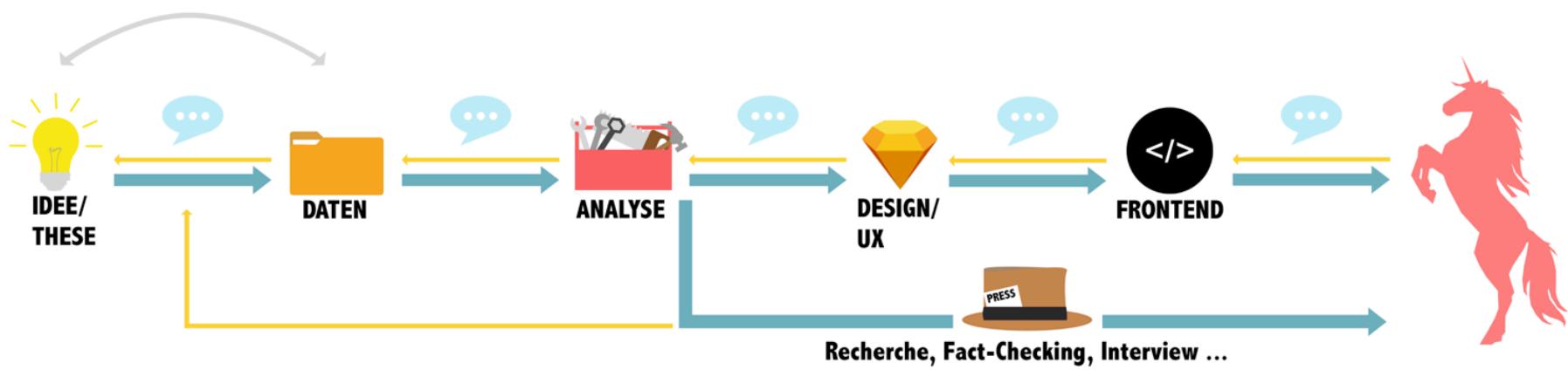
Data Journalism

data journalism <> data driven journalism

journalism with and about data

research method and form of journalistic presentation

Workflow



Data Journalism

Berliner Morgenpost

Gefällt mir 0



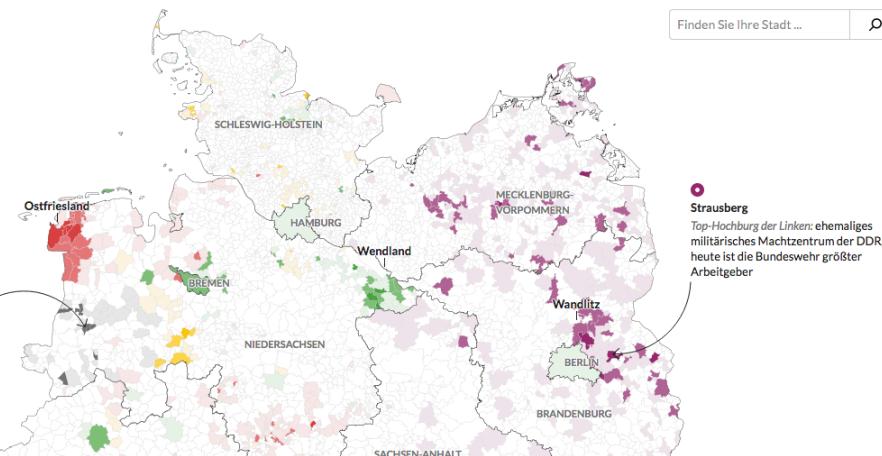
DAUER-HOCHBURGEN SEIT 1990

Wo sich die Parteien auf ihre Wähler verlassen können

Die Auswertung aller Bundestagswahlen seit 1990 zeigt, in welchen Gemeinden die Parteien immer ihre stärksten Ergebnisse geholt haben - und welche Faktoren diese Hochburgen ausmachen.

CDU/CSU SPD LINKE GRÜNE FDP
Top-10-Hochburgen
Top-100-Hochburgen
Überdurchschnittliches Ergebnis

Finden Sie Ihre Stadt ...



CDU / CSU

C wie christlich: In den Hochburgen der Union sind besonders viele Menschen Mitglieder der beiden großen christlichen Kirchen.



Ländlichkeit
Dorf, Kleinstadt, Mittelgroße Stadt, Großstadt

Dorf 74 26 0 0 Großstadt

Einkommen
1829 €
1345 1755 3451

Arbeitslose
3 %
1,4 6,7 15,4

Kirchenmitglieder
91,1 %
0,7 59 99

Ausländer
6,7 %
1 9,3 32,3

Alter
43,2 Jahre
39,9 43,9 49,6

Einbrüche
62,5 pro 100.000
14 159 697

Data Journalism

Süddeutsche Zeitung

PANAMA PAPERS

The secrets of dirty money



[German Version >](#)



Data Journalism & R

Why R?

R: project for statistical computing

Open Source

easy to learn

many of us data journalists know R

Data Journalism & R-Markdown

- Transparency and Reproducibility
- → documentation
- → source/evidence
- → information for the reader

How many people will open an R script?

How many people will be scared away by the **nerdy cody thingy**?

```
33 # 2016
34 elic_2016_1 <- read.xlsx(file = "input/elic_2016_1.xlsx", sheetIndex = 1, colClasses = "character", stringsAsFactors = F)
35 elic_2016_1 <- elic_2016_1 %>%
36   mutate(Datum = as.Date("2016-01-01 00:00:01")) %>%
37   mutate(Quartal = paste(format(Datum, "%y"), sprintf("%02i", (as.POSIXlt(Datum)$mon) %% 3L + 1L), sep="/")) %>%
38   select(GN = Geschäftsnr., Datum, Quartal, Land = Bestimmungsland, Wert = Wert..CHF., Position...Güterart = Güterart, Position..EKN = Exportkontrollnummer..EKN., Art =
39 Richtung) %>%
40   filter(Art == "Ausfuhr") %>%
41   select(-Art)
42
43 elic_2016_2 <- read.xlsx(file = "input/elic_2016_2.xlsx", sheetIndex = 1, colClasses = "character", stringsAsFactors = F)
44 elic_2016_2 <- elic_2016_2 %>%
45   mutate(Datum = as.Date("2016-04-01 00:00:01")) %>%
46   mutate(Quartal = paste(format(Datum, "%y"), sprintf("%02i", (as.POSIXlt(Datum)$mon) %% 3L + 1L), sep="/")) %>%
47   select(GN = Geschäftsnr., Datum, Quartal, Land = Bestimmungsland, Wert = Wert..CHF., Position...Güterart = Güterart, Position..EKN = Exportkontrollnummer..EKN., Art =
48 Richtung) %>%
49   filter(Art == "Ausfuhr") %>%
50   select(-Art)
51
52 elic_2016_3 <- read.xlsx(file = "input/elic_2016_3.xlsx", sheetIndex = 1, colClasses = "character", stringsAsFactors = F)
53 elic_2016_3 <- elic_2016_3 %>%
54   mutate(Datum = as.Date("2016-07-01 00:00:01")) %>%
55   mutate(Quartal = paste(format(Datum, "%y"), sprintf("%02i", (as.POSIXlt(Datum)$mon) %% 3L + 1L), sep="/")) %>%
    select(GN = Geschäftsnr., Datum, Quartal, Land = Bestimmungsland, Wert = Wert..CHF., Position...Güterart = Güterart, Position..EKN = Exportkontrollnummer..EKN., Art =
Richtung) %>%
    filter(Art == "Ausfuhr") %>%
```

< versus >

Wie viele Einträge haben in der Tracker-Applikation zwei Signaturen?

```
dim(filter(tracker_summarized, NSGII..GKV. != "", WA..GKV. != ""))[1]
```

Wie viele davon sind solche, die mit der gleichen Obersignatur (sprich: dem gleichen Haupt- und Untertyp, siehe unten) beginnen?

```
dim(filter(tracker_summarized, NSGII..GKV. != "", WA..GKV. != "", substr(NSGII..GKV.,1,2) == substr(WA..GKV.,1,2)))[1]
```

Data Journalism & R-Markdown

- + PDF-like appearance (not as “scary” as code)
- + well structured with highlighted headings, subtitles etc.
- + clear cut between data description and code
- + code part still may look scary, but additional information helps to understand the analysis
- detailed documentation takes time

Data Journalism & R-Markdown

“Is the documentation really necessary?”

“Wouldn’t source citation/the raw data be enough?”

Data Journalism & R-Markdown

document like a data journalist

- name the data source
- explain the raw data (columns, e.g. as a table)
- structure different topics in chunks
- use subtitles and descriptions for the chunks
- explain what you do and why
- explain your results

Data Journalism & R-Markdown

document like a data journalist

- who's the person you address? How can you explain the analysis well to him/her?
- Keep it short! Explain what's necessary to understand and reproduce your work, but don't confuse with too many details!

Data Journalism & R-Markdown

document like a data journalist

Your turn!

- Write a good markdown for analysis.R
- You can even simplify or prettify the code to make it less confusing
- More ideas what you could get out of the data? Continue the analysis and document it!

Bye bye, R-Ladies Berlin!

