

Automated Reporting

Using R, AWS and Datawrapper to
Automatically Create Tables and Charts

Yannik Buhl | Data Journalist & Economics Editor

Twitter: @YBuhl

Stuttgart, September 23rd, 2020

Challenge

Real-time weather reporting is gaining importance due to global heating

Increasing attention from our readers for weather extremes such as droughts

What do we want?

- Tables and charts that display current weather extremes (e.g., highest temperature, longest time without rain)
- Be able to re-use and re-implement the graphics w/o thinking about them too much

What do we need?

- Reliable, frequently updated weather **data**
- A tool for automatically updating **interactive graphics**
- A tool for **automation**

What is the desired outcome?

- Example reporting
- What is the driest place (drought reporting)?
- What is the hottest place *right now*?
- What is the windiest place *right now*?
- Where did we see the highest precipitation recently?

The pipeline

- Setting up an **Amazon Web Services (AWS)** instance with **R and RStudio Server**
- Writing a R script that gets the data, processes them and sends them to the chart
- Automating the script using a **cronjob** on AWS

Introducing `{rdwd}`¹

Two main functions: `selectDWD()` and `dataDWD()`

```
download <- rdwd::selectDWD(id = temp,
                           res = "10_minutes",
                           outvec = TRUE,
                           var = "air_temperature",
                           per = "now")

res <- rdwd::dataDWD(file = download,
                    dir = here::here("wetter"),
                    force = TRUE,
                    quiet = TRUE,
                    overwrite = TRUE)
```

They basically access DWD's Open Data Portal: opendata.dwd.de/

¹ <https://github.com/brry/rdwd>

Introducing *Datawrapper*

- *Datawrapper* is a (mostly) free tool from Germany to visualise all sorts of data
- Made for news outlets, but everyone can use the free plan
- Can be embedded in HTML via `<iframe>`
- Here's an [Example](#) from our coronavirus reporting

Introducing {DatawrappR} ¹

- Relatively new R package for *Datawrapper*
- Accesses *Datawrapper*'s API (API key needed)
- Allows you to manipulate your tables and charts from within R

```
dw_data_to_chart(x = results2, chart_id = "lxp0i")
```

- Possible manipulations: Create chart, create map, update data, edit chart, export chart, ...

¹ <https://github.com/munichrocker/DatawrappR>

AWS with R & RStudio Server

- Provides you the RStudio IDE on AWS
- I have an EC2 instance ([T2.Micro, 1 GB RAM](#))
- First year is for free, now it's ~ \$12 per month
- Link to the tutorial [here](#)

The *even more precise* pipeline

1. Set up AWS instance with R and RStudio Server
2. Provide a list with DWD station IDs (in an Excel file on AWS)
3. Write R script using `{rdwd}` that downloads and processes the data
4. Use `dw_data_to_chart()` to get new data to your chart
5. Program *cronjob* to run your script every 10-15 minutes
6. Make sure to use a log file to trace errors

```
/15 * * * * /usr/bin/Rscript  
--no-save --no-restore --no-site-file --no-init-file  
/home/username/R/script.R >> /home/username/R/script.log
```