

```
library(RSQLite)
library(dbmisc)
library(dplyr)
db = dbConnect(RSQLite::SQLite(),"articles.sqlite") %>%
  set.db.schemas(schema.file=system.file("schema/articles.yaml",
package="EconJournalData"))

articles = dbGet(db,"article")
fs = dbGet(db,"files_summary")
```

Let us look grouped by journal at the share of articles whose code supplement has R files:

```
fs %>%
  left_join(select(articles, id, journ), by="id") %>%
  group_by(journ) %>%
  mutate(num_art = n_distinct(id)) %>%
  filter(file_type=="r") %>%
  summarize(
    num_art = first(num_art),
    num_with_r = n(),
    share_with_r=round((num_with_r / first(num_art))*100,2)
  ) %>%
  arrange(desc(share_with_r))
```

journ	num_art	num_with_r	share_with_r
ecta	144	19	13.19
aeri	28	3	10.71
jep	127	12	9.45
restud	312	22	7.05
jpe	155	9	5.81
aejmic	129	5	3.88
aejpol	426	15	3.52
aer	1540	53	3.44
jeea	154	5	3.25
aejapp	430	13	3.02
aejmac	314	8	2.55
restat	813	6	0.74

We see that there is quite some variation in the share of articles with R code going from 13.2% in Econometrica (ecta) to only 0.74% in the Review of Economics and Statistics (restat). (The statistics exclude all articles that don't have a code supplement or a supplement whose file types I did not analyse, e.g. because it is too large or the ZIP files are nested too deeply.)

Overall, we still have a clear dominance of Stata in economics:

```
# Number of articles with analysed data & code supplementary
n_art = n_distinct(fs$id)

# Count articles by file types and compute shares
fs %>% group_by(file_type) %>%
  summarize(count = n(), share=round((count / n_art)*100,2)) %>%
  # note that all file extensions are stored in lower case
  filter(file_type %in% c("do","r","py","jl","m")) %>%
  arrange(desc(share))
```

file_type count share

do	3338	70.44
m	1195	25.22
r	170	3.59
py	68	1.43
jl	8	0.17

Roughly 70% of the articles have Stata `do` files and a quarter Matlab `m` files and only 3.6% R files.

While R, Python and Julia increased their share over recent years, it seems not like a very strong trend yet.

```
sum_dat = fs %>%
  left_join(select(articles, year, id), by="id") %>%
  group_by(year) %>%
  mutate(n_art_year = n()) %>%
  group_by(year, file_type) %>%
  summarize(
    count = n(),
    share=round((count / first(n_art_year))*100,2)
  ) %>%
  filter(file_type %in% c("do", "r", "py", "jl", "m")) %>%
  arrange(year, desc(share))

library(ggplot2)
ggplot(sum_dat, aes(x=year, y=share, color=file_type)) +
  geom_line(size=1.5) + scale_y_log10() + theme_bw()
```

I also have a log file that anonymously stores data about which articles that have been clicked on. The code below shows the 20 most clicked on articles so far:

```
dat = read.csv("article_click.csv")

dat %>%
  group_by(article) %>%
  summarize(count=n()) %>%
  na.omit %>%
  arrange(desc(count)) %>%
  print(n=20)

## # A tibble: 2,707 x 2
##   article
##   count
## 1 Consumer Spending during Unemployment: Positive and Normative Implicat~
##   50
## 2 Do Expert Reviews Affect the Demand for Wine?
##   44
## 3 Tax Evasion and Inequality
##   38
## 4 A Macroeconomic Model of Price Swings in the Housing Market
##   35
## 5 Is Your Lawyer a Lemon? Incentives and Selection in the Public Provisi~
##   33
## 6 The Welfare Effects of Social Media
##   31
## 7 The Rise of Market Power and the Macroeconomic Implications
```

```

29
## 8 Carbon Taxes and CO2 Emissions: Sweden as a Case Study
27
## 9 Public Debt and Low Interest Rates
27
## 10 The Sad Truth about Happiness Scales
25
## 11 Job Polarization and Jobless Recoveries
24
## 12 The New Tools of Monetary Policy
24
## 13 Alcohol and Self-Control: A Field Experiment in India
23
## 14 Disease and Gender Gaps in Human Capital Investment: Evidence from Nig~
23
## 15 Some Causal Effects of an Industrial Policy
23
## 16 Food Deserts and the Causes of Nutritional Inequality
22
## 17 Minimum Wage and Real Wage Inequality: Evidence from Pass-Through to R~
22
## 18 The Cost of Reducing Greenhouse Gas Emissions
22
## 19 Adaptation to Climate Change: Evidence from US Agriculture
21
## 20 Do Parents Value School Effectiveness?
21
## # ... with 2,687 more rows

```

So far there were over 11000 thousand clicks in total. Well, that is almost twice as much as the average number of Google searches in 100 milliseconds...