The main data for app can be downloaded as a [zipped SQLite database](http://econ.mathematik.uni-ulm.de/ejd/articles.zip) from my server. Let us do some analysis.

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 | 109 | 17 | 15.6 |
| jep | 113 | 11 | 9.73 |
| restud | 216 | 12 | 5.56 |
| aejmic | 114 | 5 | 4.39 |
| aer | 1453 | 46 | 3.17 |
| aejpol | 378 | 11 | 2.91 |
| aejapp | 385 | 11 | 2.86 |
| aejmac | 282 | 8 | 2.84 |
| jeea | 115 | 2 | 1.74 |
| restat | 733 | 6 | 0.82 |

We see that there is quite some variation in the share of articles with R code going from 15.6% in Econometrica (ecta) to only 0.82% 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 analyes 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 | 2834 | 70.18 |
| m | 979 | 24.24 |
| r | 129 | 3.19 |
| py | 42 | 1.04 |
| jl | 2 | 0.05 |

Roughly 70% of the articles have Stata do files and almost a quarter Matlab m files and only slightly above 3% R files.

I also meanwhile have added a log file to the app 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: 699 x 2

## article count

##

## 1 A Macroeconomic Model of Price Swings in the Housing Market 27

## 2 Job Polarization and Jobless Recoveries 20

## 3 Tax Evasion and Inequality 19

## 4 Public Debt and Low Interest Rates 16

## 5 An Empirical Model of Tax Convexity and Self-Employment 13

## 6 Alcohol and Self-Control: A Field Experiment in India 11

## 7 Drug Innovations and Welfare Measures Computed from Market Deman~ 11

## 8 Food Deserts and the Causes of Nutritional Inequality 11

## 9 Some Causal Effects of an Industrial Policy 11

## 10 Costs Demand and Producer Price Changes 10

## 11 Breaking Bad: Mechanisms of Social Influence and the Path to Cri~ 9

## 12 Government Involvement in the Corporate Governance of Banks 8

## 13 Performance in Mixed-sex and Single-sex Tournaments: What We Can~ 8

## 14 Disease and Gender Gaps in Human Capital Investment: Evidence fr~ 7

## 15 Housing Constraints and Spatial Misallocation 7

## 16 Inherited Control and Firm Performance 7

## 17 Labor Supply and the Value of Non-work Time: Experimental Estima~ 7

## 18 Pricing in the Market for Anticancer Drugs 7

## 19 The Arrival of Fast Internet and Employment in Africa 7

## 20 The Economic Benefits of Pharmaceutical Innovations: The Case of~ 7

## # ... with 679 more rows

