XXX

R code

Arnald Puy

Contents

```r
# PRELIMINARY FUNCTIONS -----------------------------------------------------

# Function to read in all required packages in one go
loadPackages <- function(x) {
  for(i in x) {
    if(!require(i, character.only = TRUE)) {
      install.packages(i, dependencies = TRUE)
      library(i, character.only = TRUE)
    }
  }
}

# Load the packages
loadPackages(c(
  "bibliometrix", "tidyverse", "data.table", "scales", "pdfsearch", "pdftools",
  "openxlsx", "cowplot", "wesanderson"))

# Create custom theme
theme_AP <- function() {
  theme_bw() +
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          legend.background = element_rect(fill = "transparent",
                                           color = NA),
          legend.key = element_rect(fill = "transparent",
                                    color = NA),
          strip.background = element_rect(fill = "white"),
          legend.margin = margin(0.5, 0.1, 0.1, 0.1),
          legend.box.margin=margin(0.2,-2,-7,-7))
}

# Set checkpoint
dir.create(".checkpoint")
library("checkpoint")

checkpoint("2022-05-11",
           R.version ="4.2.0",
           checkpointLocation = getwd())
# VECTOR WITH NAME OF MODELS -----------------------------------------------

models <- c("WaterGAP", "PCR-GLOBWB", "MATSIRO", "H08", "JULES-W1", "MPI-HM",
            "MHM", "LPJmL", "CWatM", "CLM", "DBHM", "ORCHIDEE")

models_vec <- paste(models, "_ref.bib", sep = "")
# BIBLIOMETRIC ANALYSIS OVER ALL BIB FILES  ---------------------------------
```

```r
output <- results <- years <- journals <-  list()
for (i in 1:length(models_vec)) {

  output[[i]] <- convert2df(file = models_vec[i],
                            dbsource = "wos",
                            format = "bibtex")

  results[[i]] <- biblioAnalysis(output[[i]], sep = ";")
  years[[i]] <- data.table(results[[i]]$Years)
  journals[[i]] <- data.table(results[[i]]$Sources) %>%
    .[, SO:= str_to_title(SO)]
}
```

```r
# PLOT ----------------------------------------------------------------

names(years) <- models
tmp <- rbindlist(years, idcol = "Model")[, .N, .(V1, Model)]

# Print total number of studies
tmp[, sum(N)]
```

```
## [1] 778
```
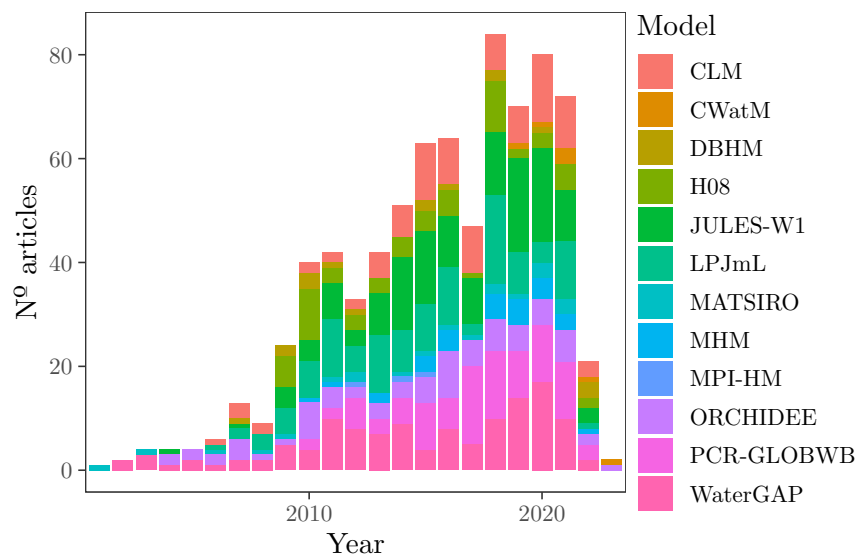
```r
plot.time <- tmp %>%
  .[, V1:= as.factor(V1)] %>%
  ggplot(., aes(V1, N, fill = Model)) +
  geom_col() +
  scale_x_discrete(breaks = pretty_breaks(n = 2)) +
  labs(x = "Year", y = "Nº articles") +
  theme_AP()

plot.time
```
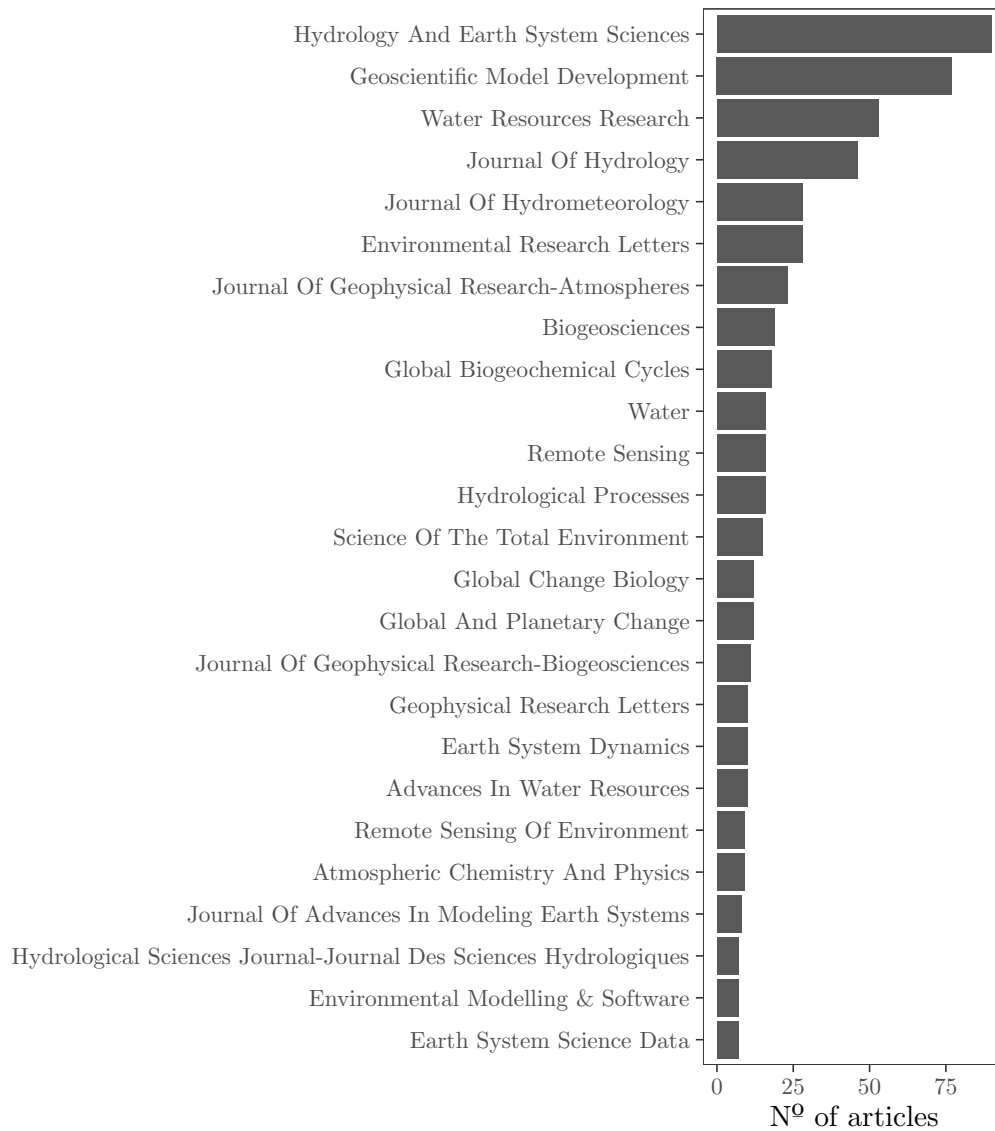
```
# PLOT JOURNALS -----------------------------------------------------------

rbindlist(journals, idcol = "Models") %>%
  .[, sum(N), SO] %>%
  .[order(-V1)] %>%
  .[, .SD[1:25]] %>%
  na.omit() %>%
  ggplot(., aes(x = reorder(SO, V1), y = V1)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(x = "", y = "Nº of articles") +
  theme_AP()
```



```
# KEYWORDS SEARCH ------------------------------------------------------

# Define vectors for search -------------------------------------------
```

```r
directory <- "/Users/arnaldpuy/Documents/papers/ghms_bibliometric/"
directory_vec <- paste(directory, models, "_pdfs", sep = "")
keywords_vec <- c("sensitivity analysis", "uncertainty analysis", "uncertainty")
filename_keywords <- paste(models, "keywords", sep = "_")

# Loop --------------------------------------------------------------------

dt <- result <- list()
for (i in 1:length(directory_vec)) {

  result[[i]] <- keyword_directory(directory_vec[i],
                                   keyword = keywords_vec,
                                   split_pdf = TRUE)

  dt[[i]] <- data.table("name" = result[[i]]$pdf_name,
                        "keyword" = result[[i]]$keyword,
                        "text" = result[[i]]$line_text)

  fwrite(dt[[i]], file = paste(filename_keywords[i], ".csv", sep = ""))

}

names(result) <- models
names(dt) <- models

# PLOT HISTOGRAMS WITH KEYWORDS -------------------------------------------

dt.keywords <- rbindlist(dt, idcol = "Model") %>%
  .[, .N, .(Model, name, keyword)] %>%
  .[, keyword:= str_to_title(keyword)]

plot.keywords.histogram <- dt.keywords %>%
  ggplot(., aes(N, fill = keyword, color = keyword)) +
  geom_histogram(position = "identity", alpha = 0.5) +
  facet_wrap(~Model, ncol = 6) +
  scale_y_continuous(breaks = pretty_breaks(n = 2)) +
  scale_x_continuous(breaks = pretty_breaks(n = 3)) +
  labs(x = "Nº of mentions", y = "Nº of papers") +
  theme_AP() +
  theme(legend.position = "top",
        strip.text.x = element_text(size = 8))

plot.keywords.histogram

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
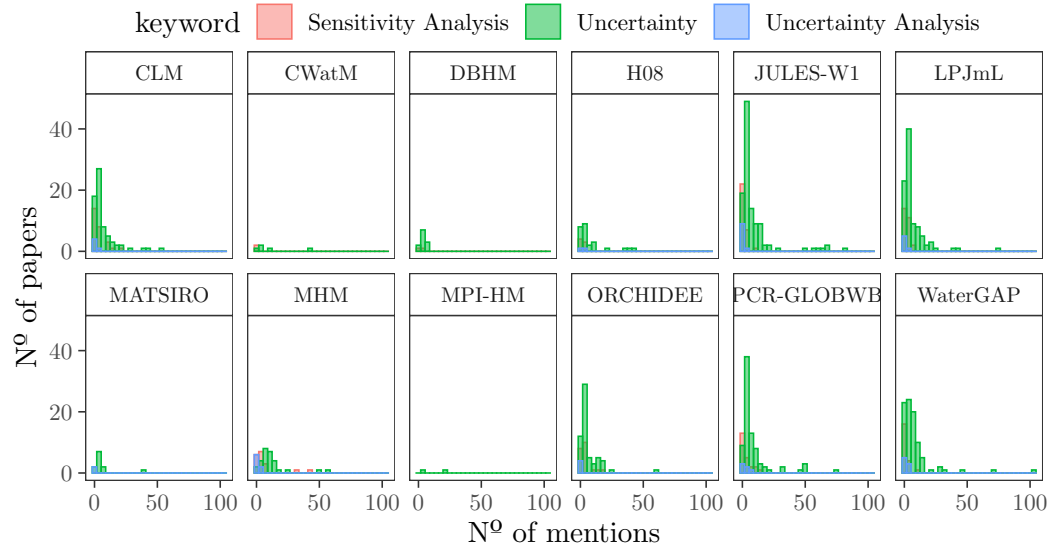
```r
# PLOT TOTAL NUMBER OF STUDIES AND TOTAL NUMBER STUDIES WITH KEYWORDS ----------

total.studies <- tmp[, sum(N), Model]
setnames(total.studies, "V1", "Total")


new.colnames <- c("N with keywords", "N Total")

dt.bars <- dt.keywords[, unique(name), Model] %>%
  .[, .N, Model] %>%
  merge(., total.studies, by = "Model") %>%
  setnames(., c("N", "Total"), new.colnames)

plot.bars <- melt(dt.bars, measure.vars = new.colnames) %>%
  ggplot(., aes(reorder(Model, value), value, fill = variable)) +
  coord_flip() +
  labs(y = "Nº of articles", x = "") +
  scale_fill_manual(values = wes_palette(2, name = "Darjeeling1"),
                    name = "") +
  geom_bar(stat = "identity", position = position_dodge(width = 0.6)) +
  theme_AP() +
  theme(legend.position = c(0.55, 0.25))

plot.bars
```
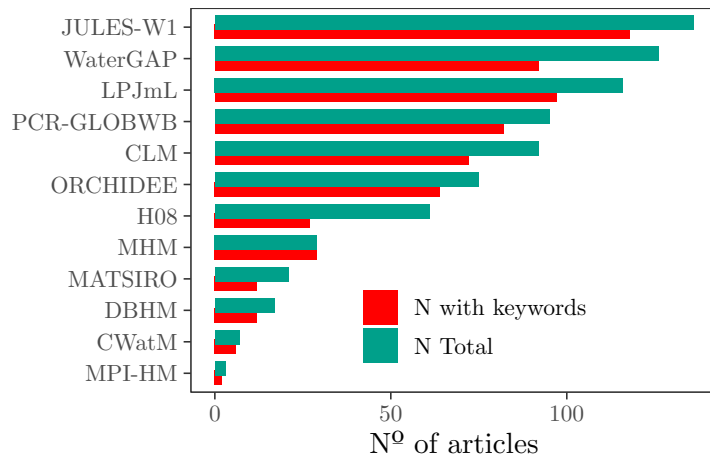
```
# MERGE PLOTS -----------------------------------------------------------------

plot_grid(plot.time, plot.bars, ncol = 2, labels = "auto", rel_widths = c(0.5, 0.5))
```