# Systematic literature review

**A papers' analysis**

Olivier Caron Christophe Benavent

2023-05-17

## Table of contents



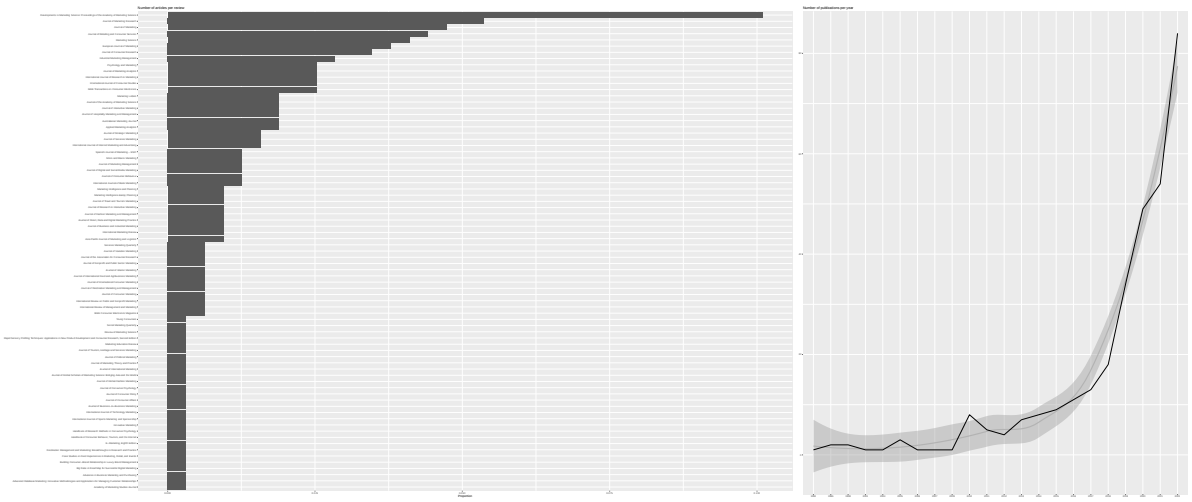
## 1 Libraries and loading data

```
library(cowplot)
library(tidyverse)
library(ggstatsplot)

#read.csv transforms ":" in "." so it was complicated to replicate code
#every column name is now lowered, the dots are replaced with _
#every underscore at the start or end of the column name is deleted

nlp_papers <- read.csv("nlp_papers_utf8.csv", fileEncoding = "UTF-8")
new_colnames <- gsub("\\.", "_", colnames(nlp_papers)) %>%
  trimws(., whitespace = "_") %>%
  gsub("_+", "_", .) %>%
  tolower()


colnames(nlp_papers) <- new_colnames
```

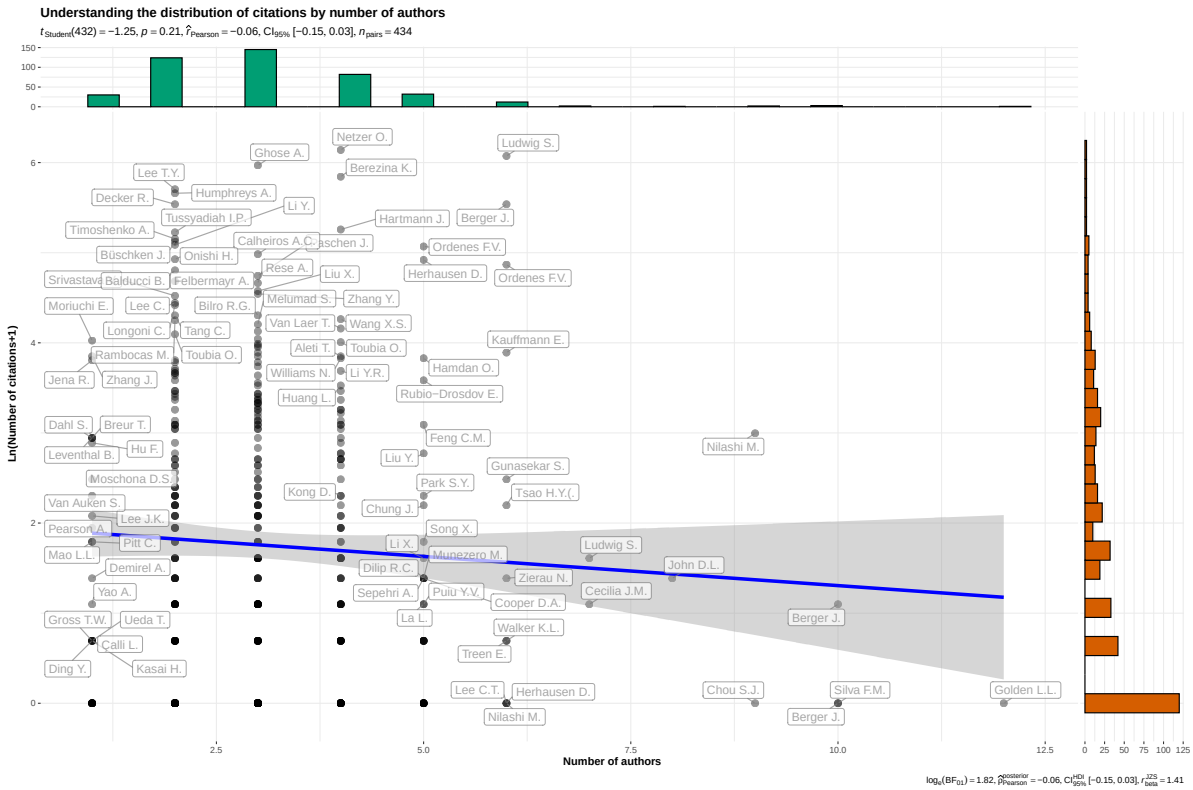
## 2 A look at the number of marketing publications with NLP

The number of papers in marketing reviews using NLP methods has increased a lot in recent years:

```
nlp_papers <- nlp_papers %>%
  mutate(year = substr(prism_coverdate,1,4))


#get rid of conference papers
nlp_papers_journal_only <- nlp_papers %>%
  filter(!grepl("conference", subtypedescription, ignore.case = TRUE)) %>%
  filter(year < 2023)



t0 <-as.data.frame(prop.table(table(nlp_papers_journal_only$prism_publicationname)))
g01<-ggplot(t0,aes(x=reorder(Var1, Freq), y=Freq))+geom_bar(stat="identity")+
  coord_flip()+
  labs( title="Number of articles per review", y="Proportion", x= NULL)

t1<-as.data.frame(table(nlp_papers_journal_only$year))

g02<-ggplot(t1, aes(x=Var1, y=Freq, group=1))+
  geom_smooth(color="Grey70", linewidth=2)+
  geom_line(stat="identity", size=1.1) +
  labs( title="Number of publications per year", y="", x=NULL)
plotgrid <- plot_grid(g01,
                      g02,
                      label_size = 10,
                      ncol=2,
                      rel_widths =  c(2,1))
ggsave(filename="images/evolution_publications_nlp_marketing.png",
       width = 80,
       height = 40,
       units = "cm")
plotgrid
```



```
nlp_papers$log_citedby = log(nlp_papers$citedby_count+1)
ggscatterstats(
  data  = nlp_papers,
  x     = author_count,
  y     = log_citedby,
  xlab  = "Number of authors",
  ylab  = "Ln(Number of citations+1)",
  title = "Understanding the distribution of citations by number of authors",
  label.var = dc_creator,
  point.label.args = list(alpha = 0.7, size = 4, color = "grey50"),
  xfill = "#CC79A7", ## fill for marginals on the x-axis
  yfill = "#009E73" ## fill for marginals on the y-axis
)
```




```
ggsave("images/distribution_citations_authors.svg", width=15, height = 12)
```