Package GetLattesData. This package is very handy  
for anyone that researches bibliometric data of Brazilian scholars. You  
could easily import the whole academic history of any researcher  
registered at the platform.

However, a couple months ago CNPQ introduced a captcha in the webpage.  
This made it impossible to download the xml files directly, breaking my  
code. It seems that those changes are now permanent. The update to  
GetLattesData will address this issue by asking the user to download the  
files manually and input its location to function  
gld\_get\_lattes\_data\_from\_zip.

Next I provide an example of usage from the vignette:

library(GetLattesData)

# get files from pkg (you can download from other researchers in lattes website)

f.in <- c(system.file('extdata/3262699324398819.zip', package = 'GetLattesData'),

system.file('extdata/8373564643000623.zip', package = 'GetLattesData'))

# set qualis

field.qualis = 'ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS, CIÊNCIAS CONTÁBEIS E TURISMO'

# get data

l.out <- gld\_get\_lattes\_data\_from\_zip(f.in,

field.qualis = field.qualis )

##

## Reading 3262699324398819.zip - Marcelo Scherer Perlin

## Found 21 published papers

## Found 2 accepted paper(s)

## Found 10 supervisions

## Found 2 published books

## Found 0 book chapters

## Found 17 conference papers

## Reading 8373564643000623.zip - Denis Borenstein

## Found 75 published papers

## Found 2 accepted paper(s)

## Found 97 supervisions

## Found 1 published books

## Found 6 book chapters

## Found 89 conference papers

The output my.l is a list with the following dataframes:

names(l.out)

## [1] "tpesq" "tpublic.published" "tpublic.accepted"

## [4] "tsupervisions" "tbooks" "tconferences"

The first is a dataframe with information about researchers:

tpesq <- l.out$tpesq

str(tpesq)

## 'data.frame': 2 obs. of 9 variables:

## $ name : chr "Marcelo Scherer Perlin" "Denis Borenstein"

## $ last.update : Date, format: "2018-09-24" "2018-08-24"

## $ phd.institution: chr "University of Reading" "University of Strathclyde"

## $ phd.start.year : num 2007 1991

## $ phd.end.year : num 2010 1995

## $ country.origin : chr "Brasil" "Brasil"

## $ major.field : chr "CIENCIAS\_SOCIAIS\_APLICADAS" "ENGENHARIAS"

## $ minor.field : chr "Administração" "Engenharia de Produção"

## $ id.file : chr "3262699324398819.zip" "8373564643000623.zip"

The second dataframe contains information about all published  
publications, including Qualis and SJR:

dplyr::glimpse(l.out$tpublic.published)

## Observations: 96

## Variables: 12

## $ id.file "3262699324398819.zip", "3262699324398819.zip", ...

## $ name "Marcelo Scherer Perlin", "Marcelo Scherer Perli...

## $ article.title "Teoria do Caos aplicada aos Contratos de Café n...

## $ year 2006, 2009, 2007, 2011, 2013, 2013, 2013, 2013, ...

## $ language "Português", "Inglês", "Inglês", "Inglês", "Port...

## $ journal.title "READ - Revista Eletrônica da Administração (UFR...

## $ ISSN "-", "1753-9641", "1413-2311", "1749-9135", "167...

## $ order.aut 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 2, 2, 1, 1, 3, 1, ...

## $ n.authors 2, 1, 2, 2, 1, 3, 3, 3, 2, 2, 3, 2, 4, 5, 3, 2, ...

## $ qualis NA, NA, "B1", NA, "B1", "A2", "B1", "A1", "B1", ...

## $ SJR NA, 0.213, NA, NA, NA, 0.886, NA, 0.429, NA, NA,...

## $ H.SJR NA, 6, NA, NA, NA, 17, NA, 38, NA, NA, NA, NA, 4...

Other dataframes in l.out included information about accepted papers,  
supervisions, books and conferences.

**An application of GetLattesData**

GetLattesData makes it easy to create academic reports for a large  
number of researchers. See next, where we plot the number of  
publications for each researcher, conditioning on Qualis ranking.

tpublic.published <- l.out$tpublic.published

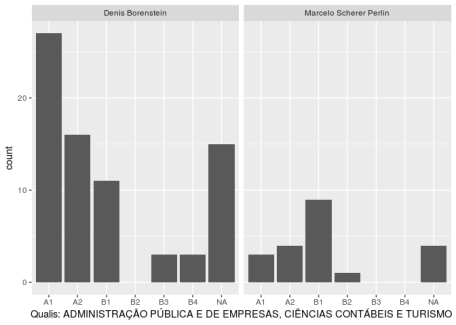
library(ggplot2)

p <- ggplot(tpublic.published, aes(x = qualis)) +

geom\_bar(position = 'identity') + facet\_wrap(~name) +

labs(x = paste0('Qualis: ', field.qualis))

print(p)



We can also use dplyr to do some simple assessment of academic  
productivity:

library(dplyr)

##

## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':

##

## filter, lag

## The following objects are masked from 'package:base':

##

## intersect, setdiff, setequal, union

my.tab <- tpublic.published %>%

group\_by(name) %>%

summarise(n.papers = n(),

max.SJR = max(SJR, na.rm = T),

mean.SJR = mean(SJR, na.rm = T),

n.A1.qualis = sum(qualis == 'A1', na.rm = T),

n.A2.qualis = sum(qualis == 'A2', na.rm = T),

median.authorship = median(as.numeric(order.aut), na.rm = T ))

knitr::kable(my.tab)

| **name** | **n.papers** | **max.SJR** | **mean.SJR** | **n.A1.qualis** | **n.A2.qualis** | **median.authorship** |
| --- | --- | --- | --- | --- | --- | --- |
| Denis Borenstein | 75 | 3.674 | 1.2808113 | 27 | 16 | 2 |
| Marcelo Scherer Perlin | 21 | 2.029 | 0.7204444 | 3 | 4 | 1 |