**Highlights of 2018**

This year the first version of GetDFPData released, a R package for accessing a large database of financial information from B3, the Brazilian exchange. I’m glad to report that many people are using it for their own research. I can see the number of visits in the web interface and the frequent emails I get about the package. The feedback from other researchers has been great but, off course, there are always ways to improve the code. I’ve been constantly developing it over time.

The GetDFPData package also had an impact in my own research. I’ve always been biased towards the topic of capital markets and now I’m doing research in corporate finance, mostly due to the new access to a large database of corporate events. Currently, I have three paper initiatives in analyzing the effect of boards formation towards financial performance of Brazilian companies.

Overall, 2018 was a great year. I’m always thankful for having the opportunity of working in a job that I love and look forward to work (almost) every single day.

**Blog posts in 2018**

First, let’s see how many posts the website has so far.

my.blog.folder <- '~/Dropbox/11-My Website/www.XYZ.com-blog/content/post/'

post.files <- list.files(path = my.blog.folder, pattern = '.Rmd')

post.files

## [1] "2017-01-01-First-post.Rmd"

## [2] "2017-01-02-GetHFData.Rmd"

## [3] "2017-01-15-CalculatingBetas.Rmd"

## [4] "2017-01-30-Exams-with-dynamic-content.Rmd"

## [5] "2017-02-13-R-and-Tennis-Players.Rmd"

## [6] "2017-02-16-Writing-a-book.Rmd"

## [7] "2017-03-05-Prophet-and\_stock-market.Rmd"

## [8] "2017-05-04-pafdR-is-out.Rmd"

## [9] "2017-05-09-Studying-Pkg-Names.Rmd"

## [10] "2017-05-15-R-Finance.Rmd"

## [11] "2017-08-12-Switching-to-Linux.Rmd"

## [12] "2017-09-14-Brazilian-Yield-Curve.Rmd"

## [13] "2017-12-06-Package-GetDFPData.Rmd"

## [14] "2017-12-13-Serving-shiny-apps-internet.Rmd"

## [15] "2017-12-30-Looking-Back-2017.Rmd"

## [16] "2018-01-22-Update-BatchGetSymbols.Rmd"

## [17] "2018-03-16-Writing\_Papers\_About\_Pkgs.Rmd"

## [18] "2018-04-22-predatory-scientometrics.Rmd"

## [19] "2018-05-12-Investing-Long-Run.Rmd"

## [20] "2018-06-12-padfR-ed2.Rmd"

## [21] "2018-06-29-BenchmarkingSSD.Rmd"

## [22] "2018-10-10-BatchGetSymbols-NewVersion.Rmd"

## [23] "2018-10-11-Update-GetLattesData.Rmd"

## [24] "2018-10-13-NewPackage-PkgsFromFiles.Rmd"

## [25] "2018-10-19-R-and-loops.Rmd"

## [26] "2018-10-20-Linux-and-R.Rmd"

## [27] "2018-11-03-NewBlog.Rmd"

## [28] "2018-11-03-RstudioTricks.Rmd"

## [29] "2019-01-08-Looking-Back-2018.Rmd"

The blog started in january 2017 and, over time, I wrote 29 posts. That feels alright. I’m not felling forced to write and I do it whenever I fell like I have something to share.

Let’s get more information from the .Rmd files. I’ll write function read\_blog\_files and use it for all post files.

read\_blog\_files <- function(f.in) {

require(tidyverse)

my.front.matter <- rmarkdown::yaml\_front\_matter(f.in)

df.out <- data\_frame(post.title = my.front.matter$title,

post.date = lubridate::ymd(my.front.matter$date),

post.month = as.Date(format(post.date, '%Y-%m-01')),

tags = paste0(my.front.matter$tags, collapse = ';'),

categories = paste0(my.front.matter$categories, collapse = ';'),

content = paste0(read\_lines(f.in), collapse = ' '))

return(df.out)

}

df.posts <- dplyr::bind\_rows(purrr::map(post.files, read\_blog\_files))

## Loading required package: tidyverse

## ── Attaching packages ────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.5

## ✔ tibble 2.0.0 ✔ dplyr 0.7.8

## ✔ tidyr 0.8.2 ✔ stringr 1.3.1

## ✔ readr 1.3.1 ✔ forcats 0.3.0

## ── Conflicts ───────────────────────────────────── tidyverse\_conflicts() ──

## ✖ dplyr::filter() masks stats::filter()

## ✖ dplyr::lag() masks stats::lag()

## Warning: `data\_frame()` is deprecated, use `tibble()`.

## This warning is displayed once per session.

glimpse(df.posts)

## Observations: 29

## Variables: 6

## $ post.title "My first post!", "Using R to download high frequency…

## $ post.date 2017-01-01, 2017-01-02, 2017-01-15, 2017-01-30, 2017…

## $ post.month 2017-01-01, 2017-01-01, 2017-01-01, 2017-01-01, 2017…

## $ tags "about me", "R;GetHFData;B3;market microstructure;hig…

## $ categories "about me", "R;GetHFData;B3;market microstructure;hig…

## $ content "--- title: \"My first post!\" subtitle: \"A little b…

First, we’ll look at the frequency of posts over time.

df.posts.2018 <- df.posts %>%

filter(post.date > as.Date('2018-01-01'))

print( ggplot(df.posts.2018, aes(x = post.month)) + geom\_histogram(stat='count') +

theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +

labs(y = 'Number of posts', x = ''))

## Warning: Ignoring unknown parameters: binwidth, bins, pad

Seems to average about once a month. The blank spaces show that I did not write for a couple of months.

**Checking 2018’s plans**

**Start a portal for financial data in Brazil**

Unfortunately this project did not launch. I wrote a couple of R scripts for fetching and saving data automatically in my server but it never became a webpage. I started to work on other projects and the website was not a priority.