

Linguagens de Programação

Alisson Rosa

5/22/2021

1 Introdução

Linguagens de programação nós permitem fornecer regras a quais os computadores entendem, ao longo do tempo foram surgindo inúmeras linguagens com propósitos e estruturas diferentes, aqui vamos analisar o comportamento dessas linguagens no GitHub (gh), que hoje é uma das principais plataformas para versionamento de códigos, assim vamos avaliar:

- Quantidade de *issues* em repositórios que contém tais linguagens ao longo do tempo
- Quantidade de *pull requests* (pr) em repositórios que contém tais linguagens ao longo do tempo
- Quantidade totais de repositórios das linguagens.

Todo código a ser desenvolvido aqui será usando a linguagem Julia.

```
include("utils.jl")
using DataFrames
import CSV
using Plots
default(formatter=identity, tickfontsize = 7, titlefontsize=12,
legend=:topleft)
#theme(:vibrant)
using Statistics
using Pipe: @pipe

pr = DataFrame(CSV.File("data/prs.csv"));
issues = DataFrame(CSV.File("data/issues.csv"));
total = DataFrame(CSV.File("data/repos.csv"));
```

Precisamos primeiro notar que até ‘ existiam linguagens disponíveis no gh, assim aqui vamos avaliar as top 5 linguagens e R¹ em quantidade de repositórios

¹R, pois certas pessoas curtem R, certo.. Certo??

Linguagem	Total de Repositórios
JavaScript	1100421
CSS	813443
HTML	779549
Shell	638068
Python	548870

Notamos que as top 3 linguagens, são de desenvolvimento web, sendo elas JavaScript, CSS, HTML e no top 4 está o famoso shell script, e por último Python. No seguinte gráfico vamos ver como se comporta R em frente a esses gigantes.

```
@pipe total |>
  sort(_, order(:num_repos, rev=true)) |>
  first(_, 5) |>
  select(_, :language => :Linguagem, :num_repos => :N) |>
  df -> bar(df.Linguagem, df.N, fill=["#f3ff33", "#337aff",
    "#ff5e33", "#55ff33", "#233cad"],
    series_annotation=df.N, legend=false)

@pipe total |>
  filter(:language ==("R"), _) |>
  df -> bar!(df[:, :language] , df[:, :num_repos], legend=false)
```

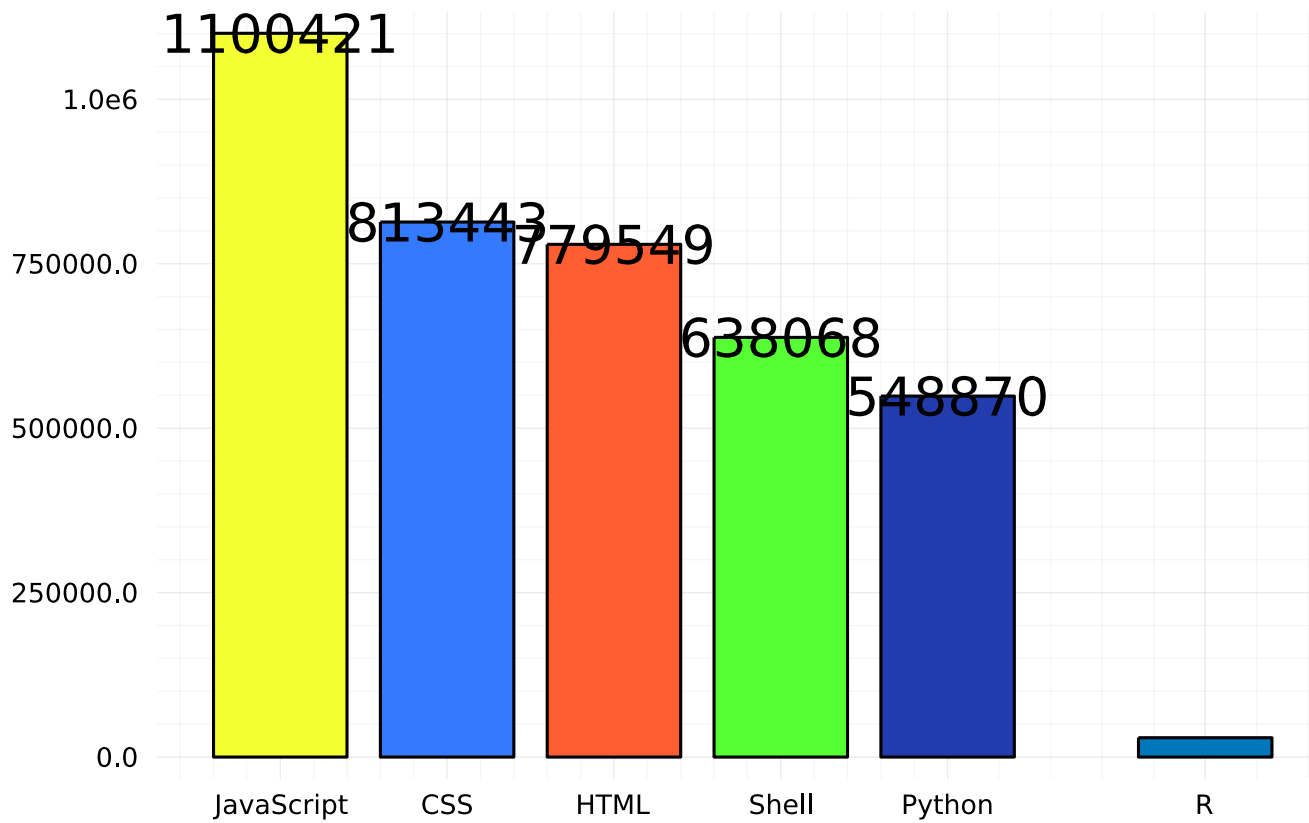


Figura 1: Top 5 Linguagens em quantidade de repositórios e R

2 Análise descritiva

Primeiro vamos começar avaliando o comportamento médio do gh ao longo dos anos

```
p1 = @pipe top_5(issues,:year) |>
  areaplot(_[:,:year],_[:,:media], title="pr", label= "prs", legend=:topleft)
p2 = @pipe top_5(pr,:year) |>
  areaplot(_[:,:year],_[:,:media], title="issues", color = "#ff4933", label= "issues",

plot(p1,p2,layout= (1,2))
```

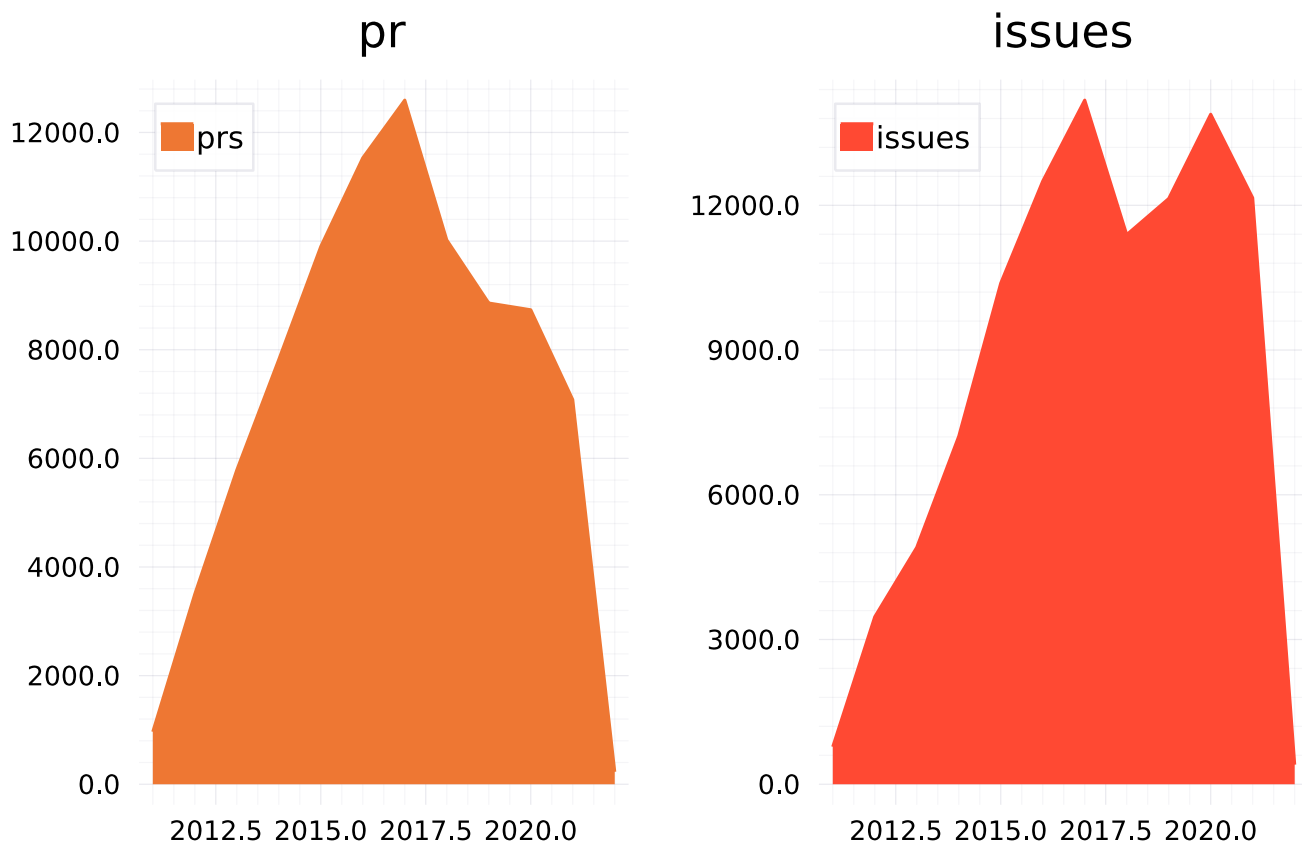


Figura 2: Comportamento médio ao longo dos anos

3 Linguagens Globalmente

Avaliar nessa seção as top 10 linguagens + R

4 Linguagens para Dados

Nessa seção vamos avaliar o comportamento especificamente de R e Julia, que são duas das três principais linguagens para trabalhar com dados no momento.

4.1 Leve introdução

R tem suas raízes na estatística, mas depois do desenvolvimento de certos pacotes se tornou um ambiente propício para ciência de dados. Julia por outro lado, não nasceu fundamentalmente para a estatística, porém como possui as características principais de ser **rápida** e **fácil** de usar, não demorou para aparecer inúmeros pacotes para análise de dados.

Vamos começar avaliando a quantidade total de repositórios para cada uma das duas linguagens.

Podemos ver que R possui muito mais repositórios disponíveis no gh, porém R é uma linguagem que está a bastante tempo no mercado, Julia possui apenas 10 anos de idade..

```
using PrettyTables
total[findall(in(["R","Julia"]), total[!,:language]), :] |>
  d -> pretty_table(d; header = ["Language", "num_repos"], backend = Val(:latex))
```

```
\begin{table}
\begin{tabular}{rr}
\hline\hline
\textbf{Language} & \textbf{num_repos} \\\hline
R & 29380 \\
Julia & 5858 \\\hline\hline
\end{tabular}
\end{table}
```

Agora vamos ver o comportamento das duas linguagens ao longo dos últimos anos

```
p1 = fav_langs(pr, ["R","Julia"]) |>
  df -> plot(df[:, :year], df[:, :media], g = df[:, :name])
p2 = fav_langs(issues, ["R","Julia"]) |>
  df -> plot(df[:, :year], df[:, :media], g = df[:, :name])

plot(p1,p2,layout=(1,2))
```

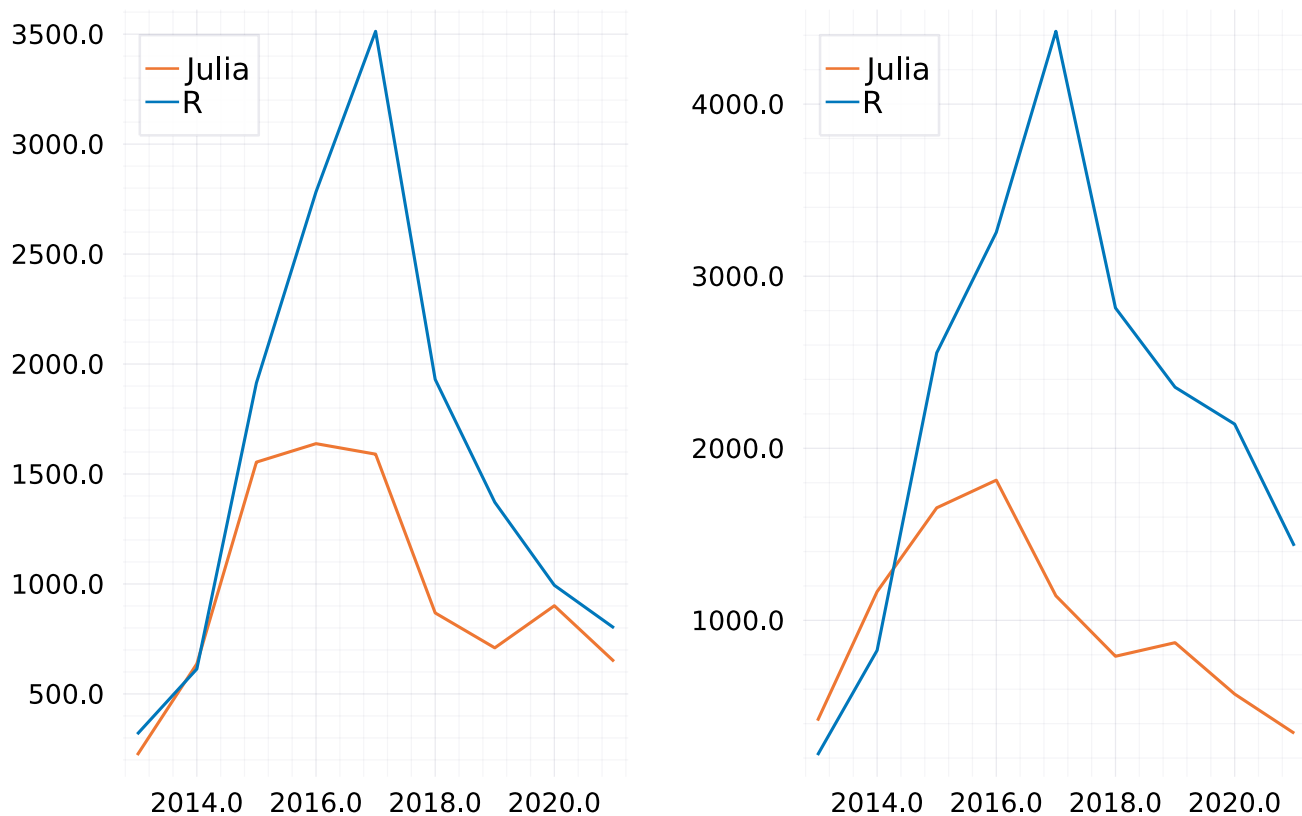


Figura 3: Média de prs ao longo dos anos

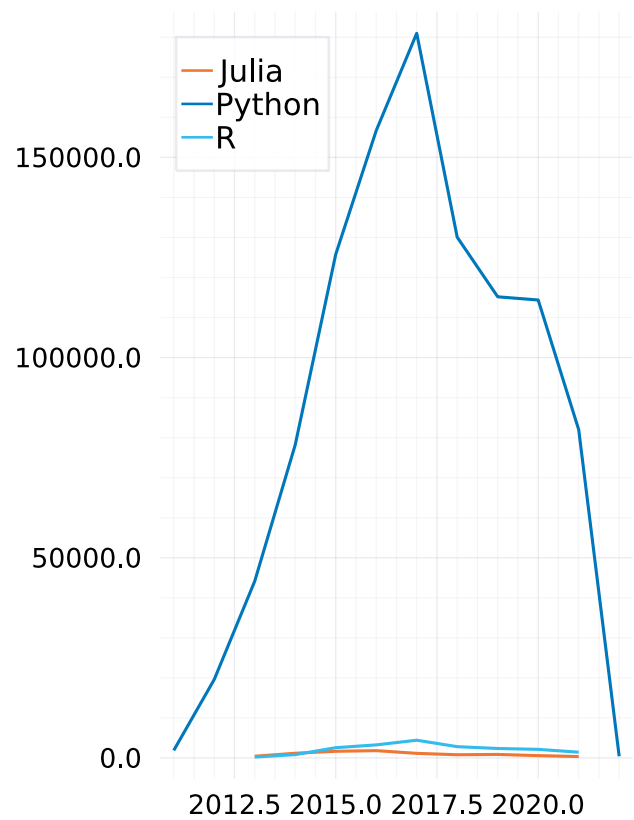
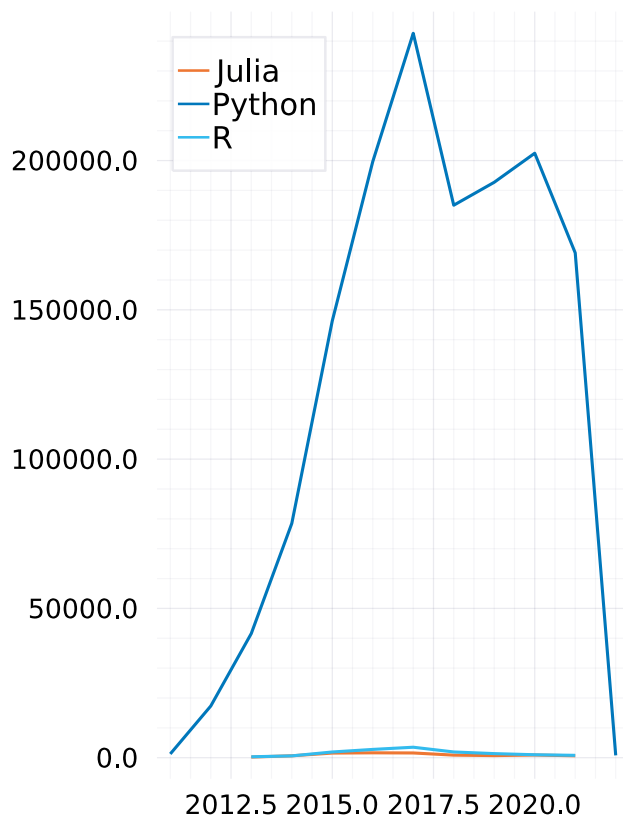
Nota-se que em 2014 Julia teve mais prs que R e ambas as linguagens atingem seu pico em 2017, um comportamento global como vimos anteriormente, porém para issues Julia teve mais de 2012 até 2014 e também seu pico de issues foi em 2016 e não 2017 como R.

4.2 Python entra no game

Evidentemente não podemos esquecer de Python, pois como vimos está no top5 de linguagens mais usadas e também ————,

```
p1 = fav_langs(pr, ["R","Julia","Python"]) |>
  df -> plot(df[:,year], df[:,media], g = df[:,name])
p2 = fav_langs(issues, ["R","Julia", "Python"]) |>
  df -> plot(df[:,year], df[:,media], g = df[:,name])

plot(p1,p2,layout=(1,2))
```



Assim pela figura notamos que: Uma imagem vale mais que mil palavras

Conclusão momentânea

Referências