

# GitHub Public Repository Metadata

Baptiste Toussaint,

## Sommaire

I	Introduction : présentation des données	3
II	Plan de recherche	8
III	Partie I : Analyse de la popularité des dépôt	10
IV	Partie II : Analyse des langages de programmation	14
V	Tableau du board	19



# I - Introduction : présentation des données



Jeu de données regroupant près de 3 millions de dépôts GitHub

Trouvé sur Kaagle, publié par Peter Elmers.

Plusieurs informations comme :

- Nom
- Date de création
- nombre d'étoiles
- etc...

```
{
  "owner": "pelmers",
  "name": "text-rewriter",
  "stars": 13,
  "forks": 5,
  "watchers": 4,
  "isFork": false,
  "isArchived": false,
  "languages": [ { "name": "JavaScript", "size": 21769 }, { "name": "HTML", "size": 2081 }, { "name": "CSS", "size": 2081 } ],
  "languageCount": 3,
  "topics": [ { "name": "chrome-extension", "stars": 43211 } ],
  "topicCount": 1,
  "diskUsageKb": 75,
  "pullRequests": 4,
  "issues": 12,
  "description": "Webextension to rewrite phrases in pages",
  "primaryLanguage": "JavaScript",
  "createdAt": "2015-03-14T22:35:11Z",
  "pushedAt": "2022-02-11T14:26:00Z",
  "defaultBranchCommitCount": 54,
  "license": null,
  "assignableUserCount": 1,
  "codeOfConduct": null,
  "forkingAllowed": true,
  "nameWithOwner": "pelmers/text-rewriter",
  "parent": null
}
```

Pourquoi ces données ?

- Nous utilisons tous GitHub, de manière différentes.
- Savoir comment les utilisateurs de GitHub se comportent.
- Un jeu de données diversifié avec beaucoup de modalités (textes, valeurs numériques, dates, etc.)



## Les données que nous avons utilisées

Nom de l'attribut	Type	Explication
<i>name</i>	char	Nom du dépôt
<i>stars</i>	int	Nombre d'étoiles
<i>forks</i>	int	Nombre de <i>forks</i>
<i>watchers</i>	int	Nombre de <i>watchers</i>
<i>primaryLanguage</i>	string	Nom du langage de programmation principalement dans le projet
<i>createdAt</i>	POSIXct	La date de création du dépôt



files-community / Files

Type to search

Issues 464

Pull requests 15

Actions

Projects 1

Security

Insights

Files Public

Sponsor

Watch 379

Fork 2.1k

Star 32.4k

main

16 Branches

114 Tags

Go to file

Add file

Code

hishitetsu

Fix: Fixed an issue where empty folders could not be compressed (#15548)

2ad172c · 6 minutes ago

5,717 Commits

.github	GitHub: Update PULL_REQUEST_TEMPLATE.md	last month
assets	[ImgBot] Optimize images (#13373)	9 months ago
builds	GitHub: Removed stable release from Azure Pipeline (#15003)	3 months ago
docs	Feature: Renamed "Favorites" section to "Pinned" (#14862)	4 months ago
scripts	GitHub: Added ps1 scripts to simplify CI/CD (#15030)	2 months ago
src	Fix: Fixed an issue where empty folders could not be compre...	6 minutes ago
tests/Files.InteractionTests	Feature: Added a setting to configure the "Open in ide" butt...	3 weeks ago
.editorconfig	Enforce tabs for Xaml in EditorConfig and Xaml Styler (#113...	last year
.gitattributes	Fixed "divide by zero" error thrown upon navigation to some...	6 years ago
.gitignore	Code Quality: Renamed Files.OpenDialog (#12845)	last year

About

Building the best file manager for Windows

files.community

git windows productivity files  
xaml csharp customization utilities  
dotnet windows-10 developer-tools  
file-manager fluent fluent-design  
alternatives power-user good-first-issue  
windows-11 winui winappsdk

Readme  
MIT license  
Code of conduct  
Activity  
Custom properties  
32.4k stars  
379 watching

## II – Plan de recherche





Deux axes d'analyse :

- La popularité des dépôts.

GitHub comme un réseau social, les étoiles, les watchers, les forks peuvent “mesurer” la popularité des dépôts.

- Les langages de programmation.

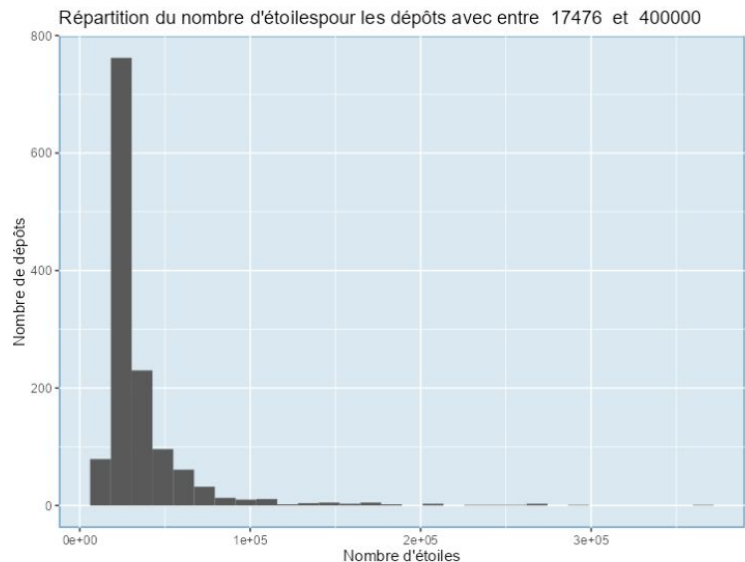
A partir de l'analyse précédente, on peut essayer de “mesurer” la popularité des langages de programmation.



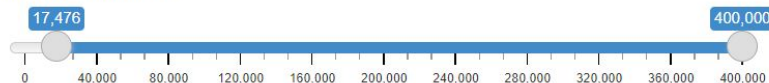
### III – Première étape : Analyse de la popularité des dépôt

### Graphiques montrant la répartition des étoiles sur les dépôts

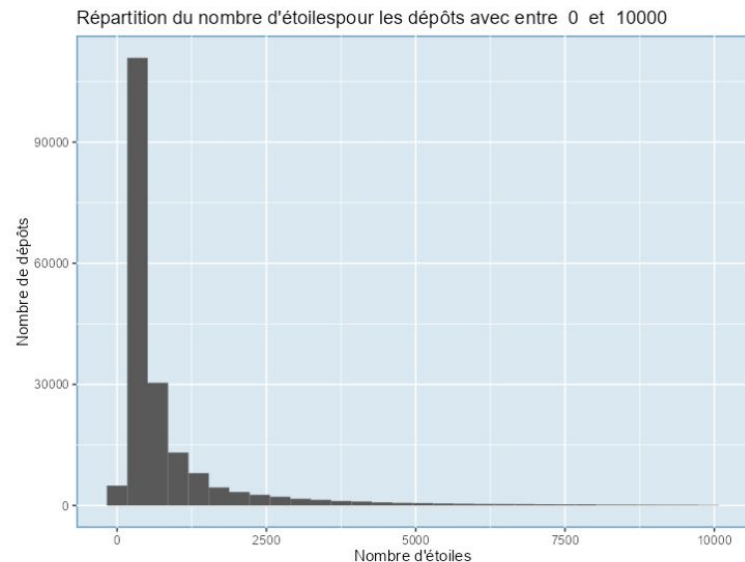
#### Toutes les données



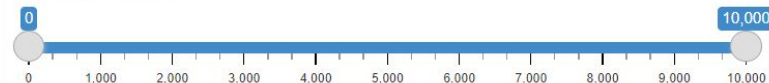
Nombre d'étoiles :



#### Seulement entre 0 et 10000 étoiles

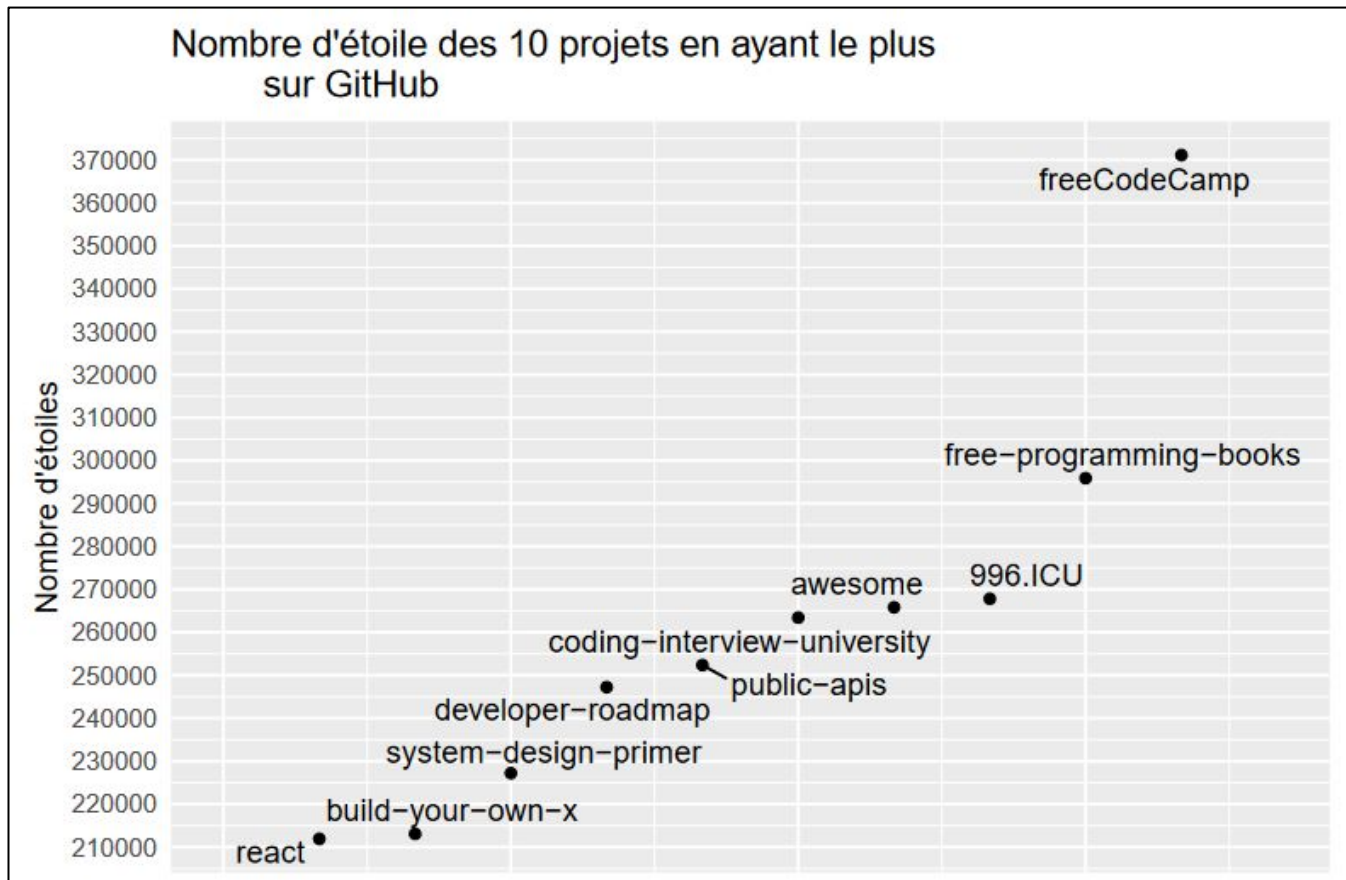


Nombre d'étoiles :



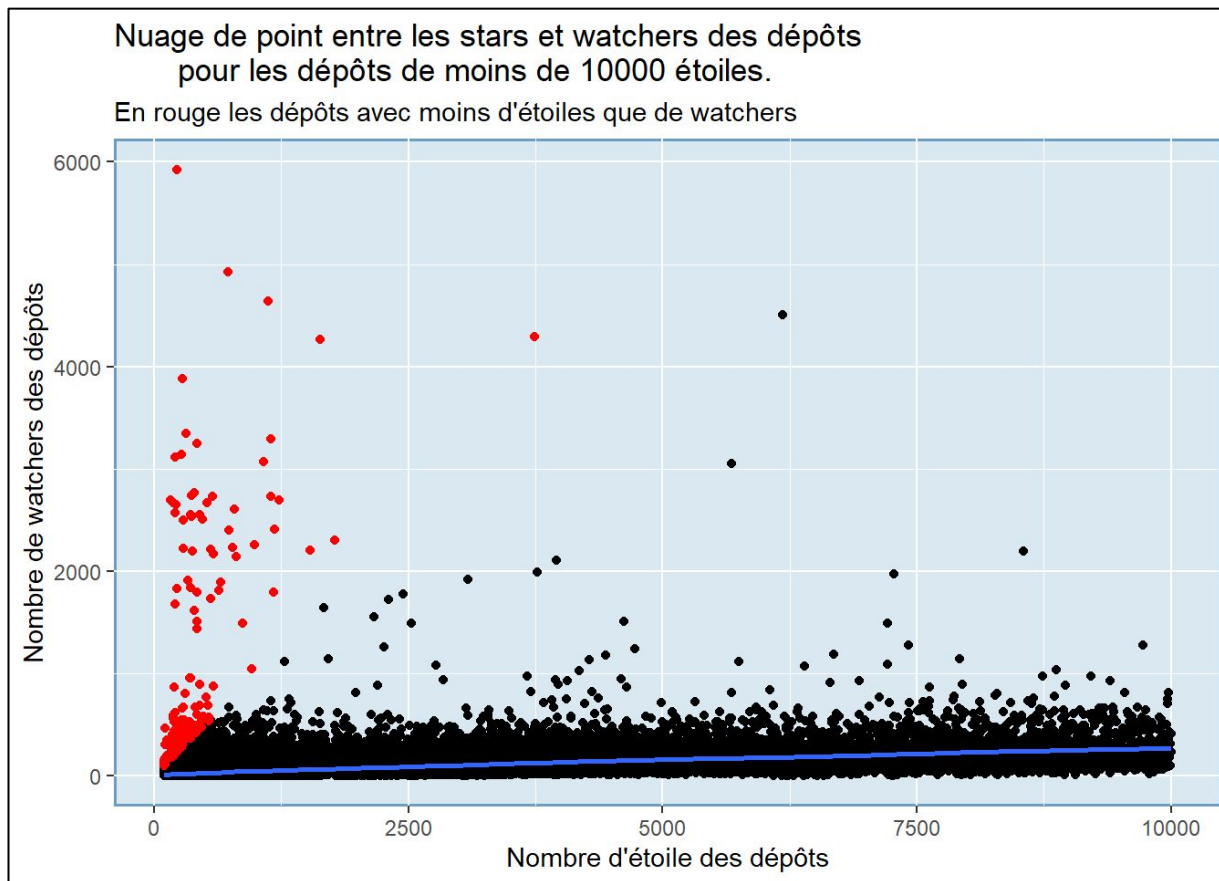
# III – Première étape : Analyse de la popularité des dépôt

Hypothèse 1 : Les dépôts très populaires sont rares

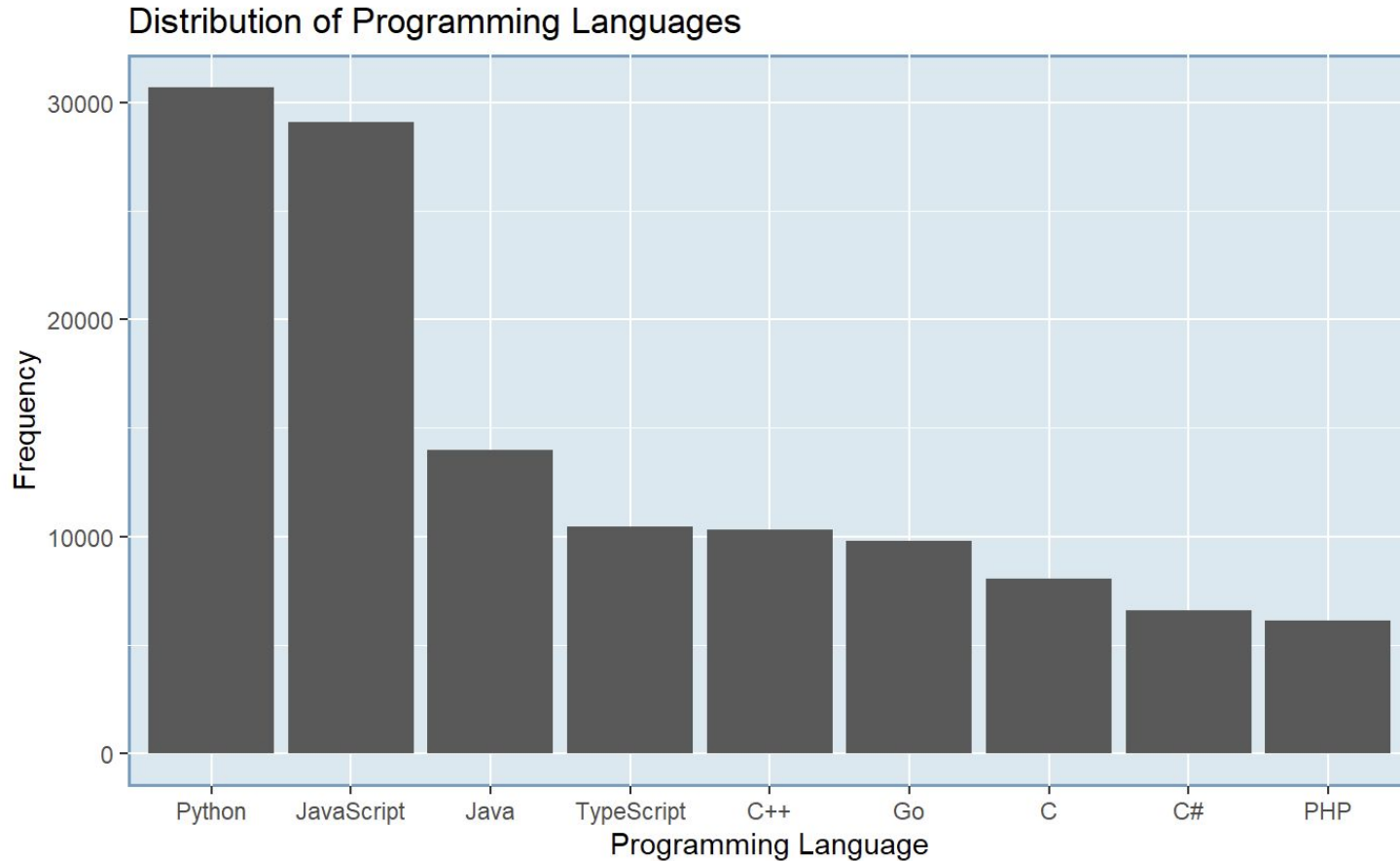


# III – Première étape : Analyse de la popularité des dépôt

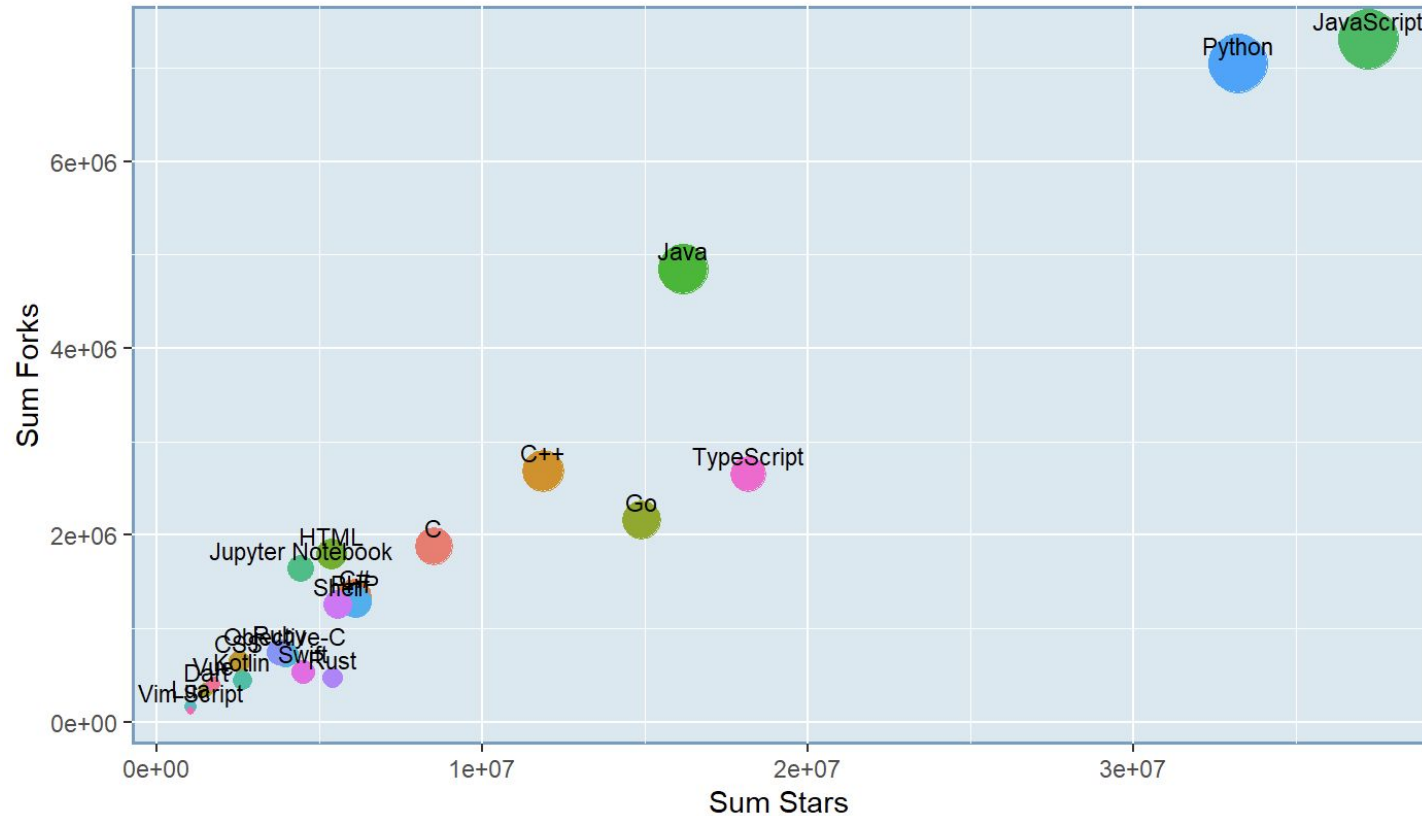
Une découverte étonnante...



## IV - Deuxième étape : Étude des langages de programmation utilisés

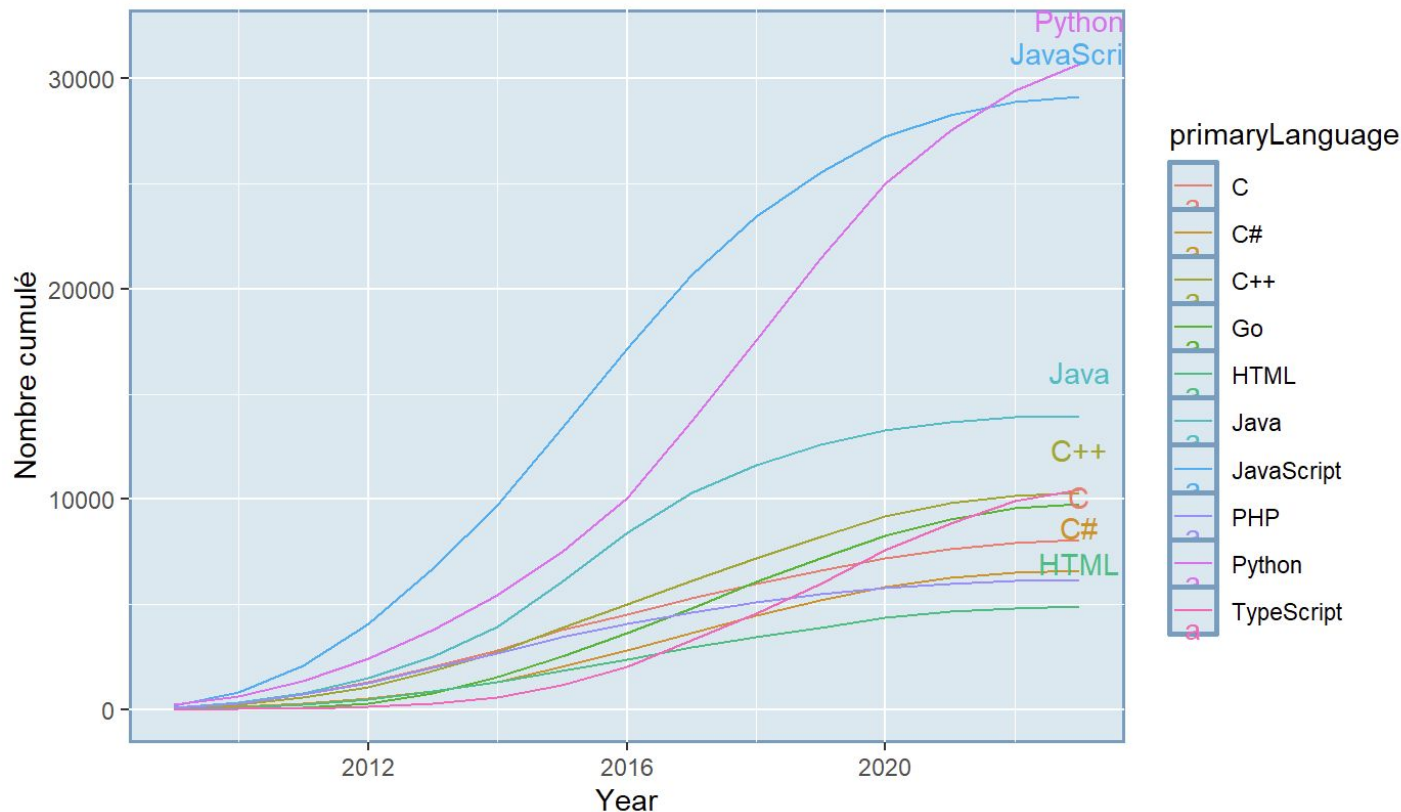


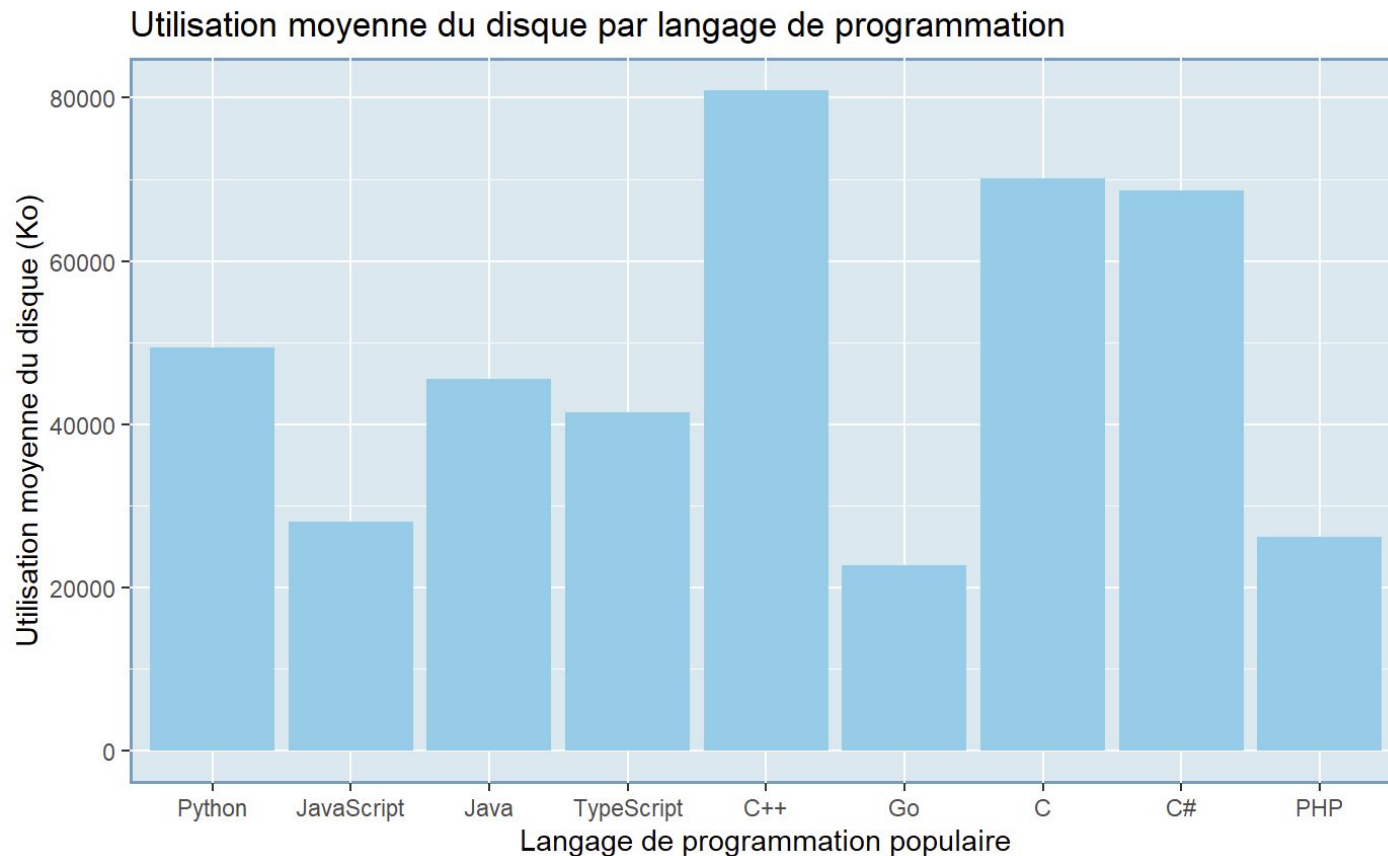
Distribution de la popularité du langage de programmation GitHub





Nombre cumulé annuel de langues principales dans les 200 000 premiers référentie

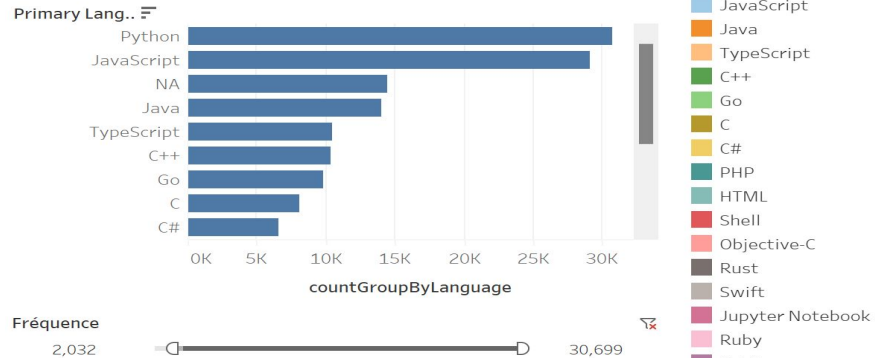




## V – Tableau du board



## La répartition des langues



## La tendance des langues

