Python Presentation

Pam, Anoop, Michael

History

Scientific Presentation :

Python in today's Present

Presenter : Anoop Johny

Semester 2

Masters Student
Life Science Informatics
Life Science Informatics Department
University of Bonn

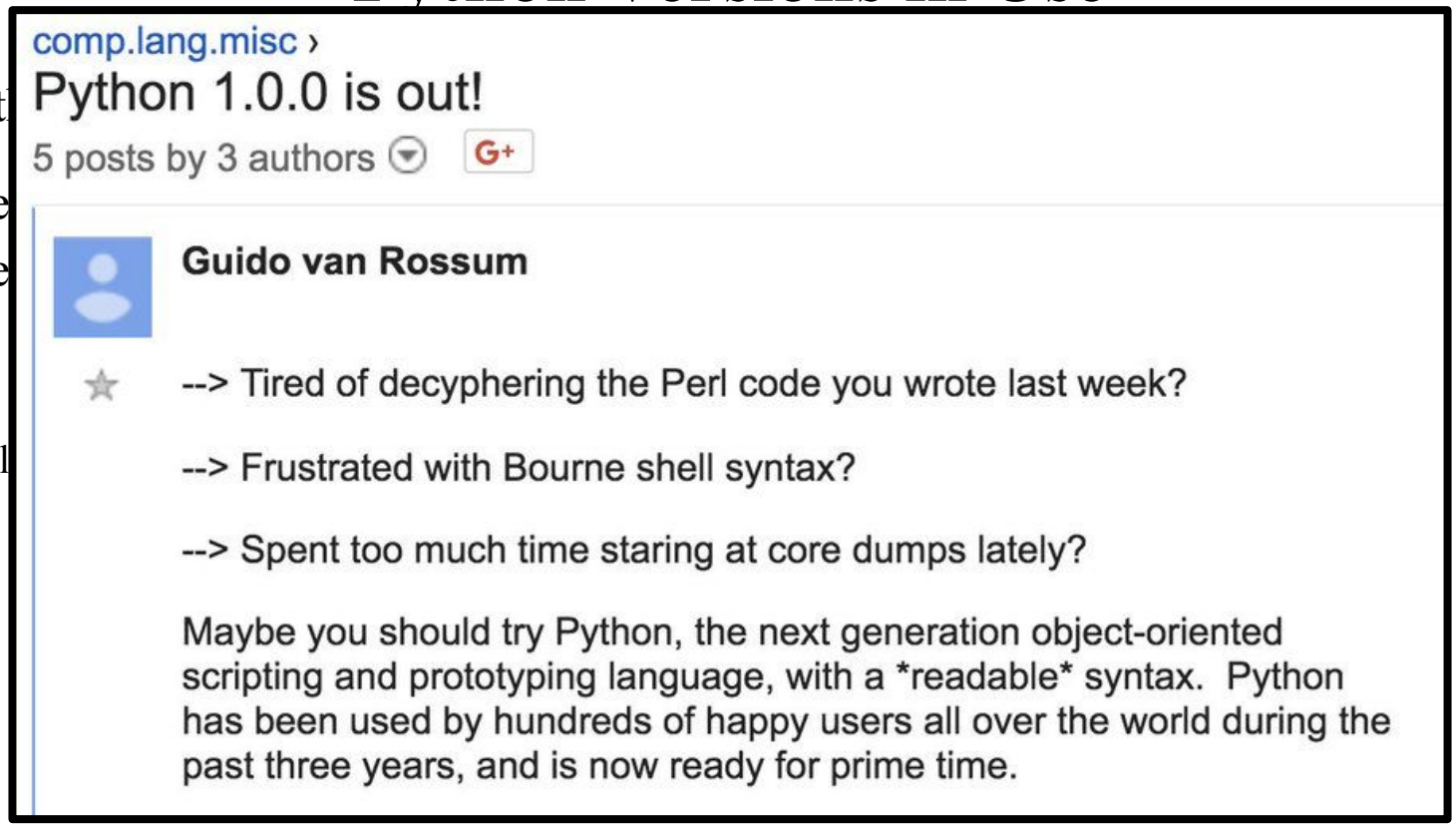
Content

Python Versions in Use

The Pyt

- the
- the

Python pl



Python Versions in Use

Original Versio

- Python 1.4

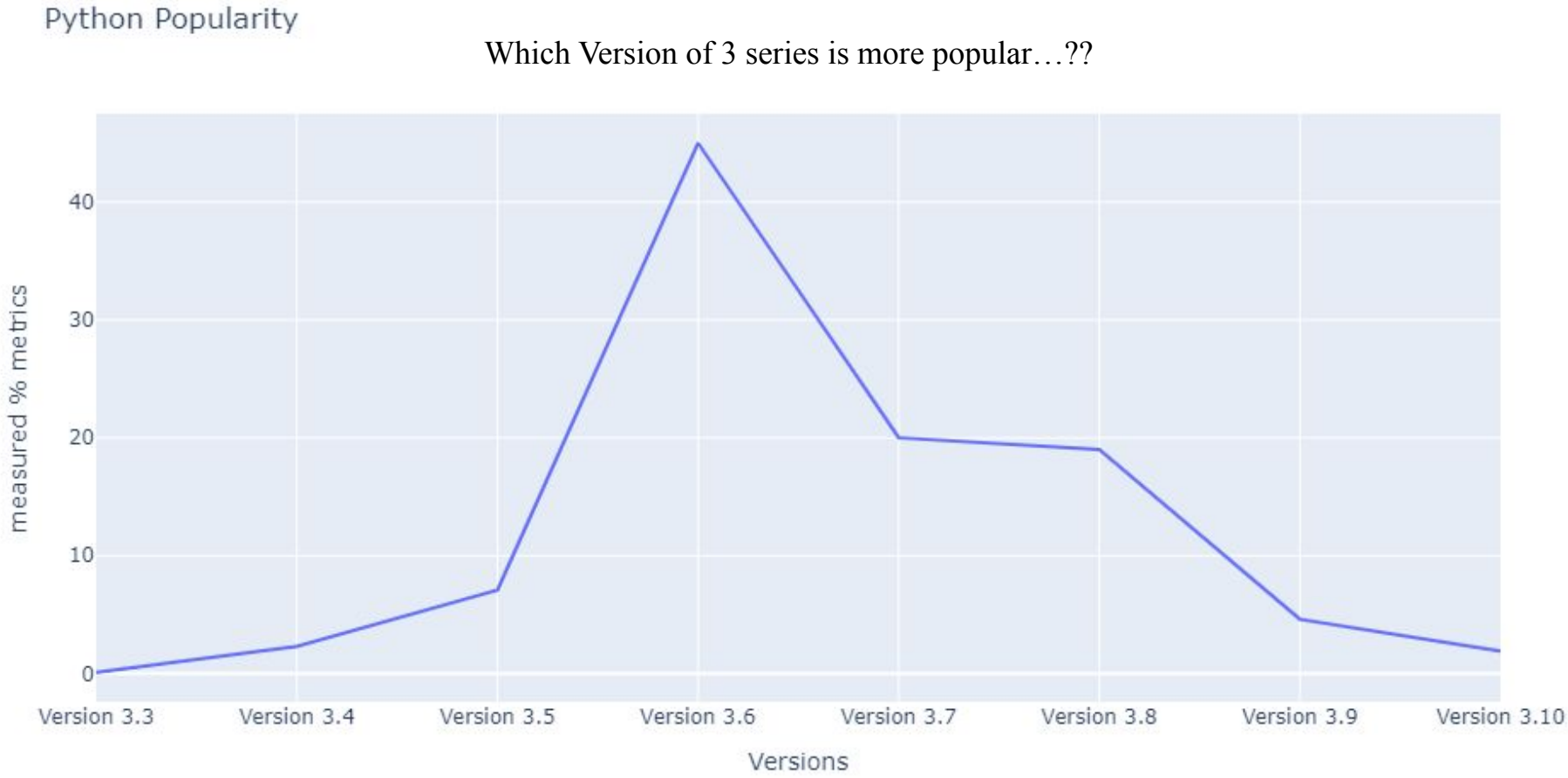
Latest Version:

- Python 3.1

Py...Use

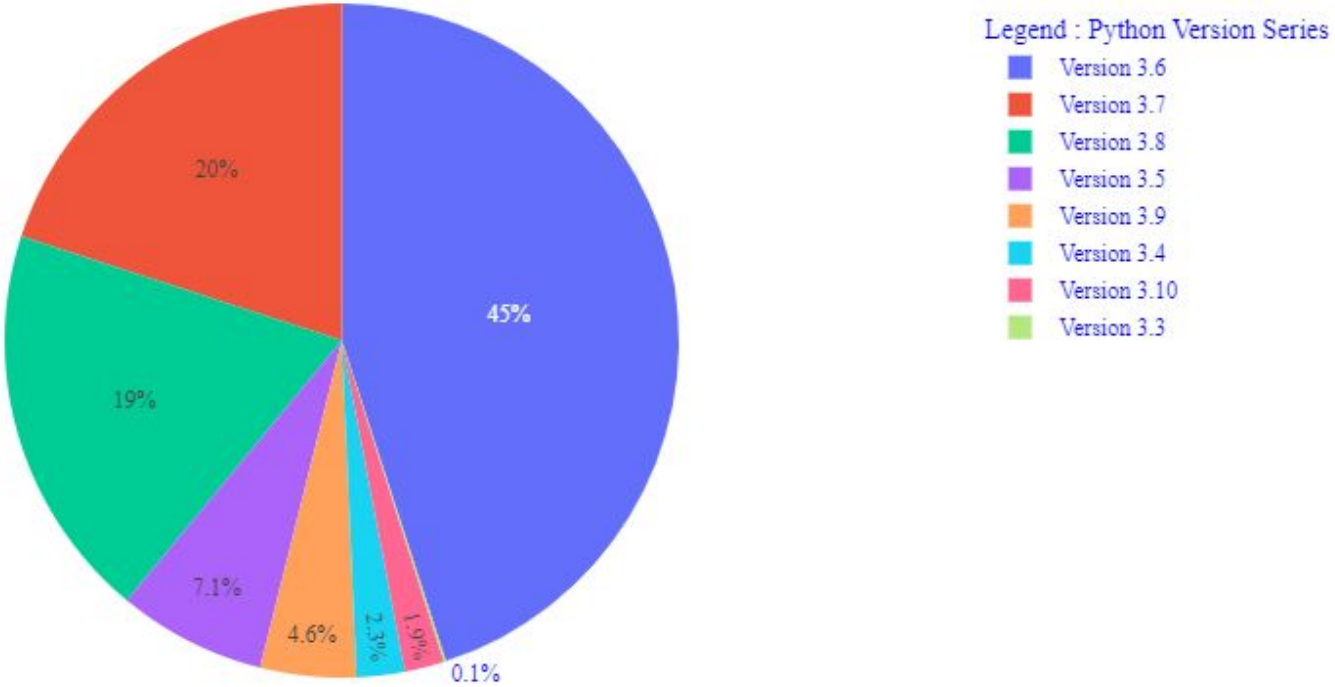
- Python Enhancement Pr...).
- Philosophy of increment...lue sky project)

Content

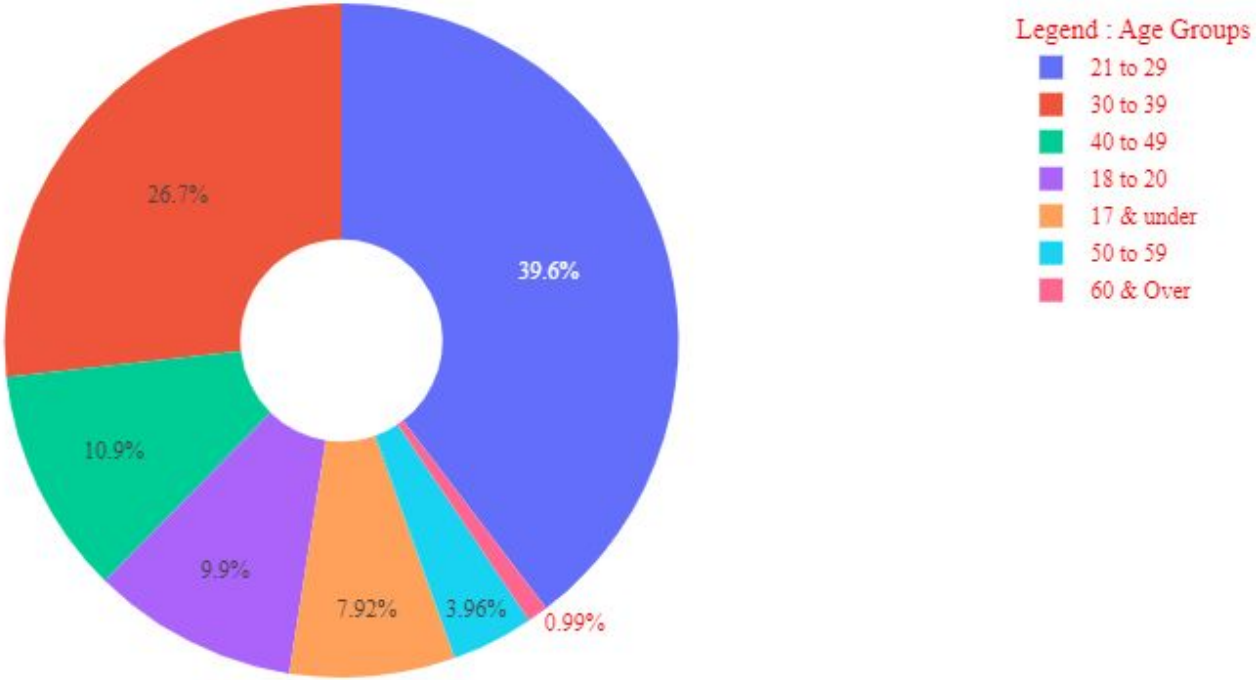

Python Popularity
Which Version of 3 series is more popular…??
measured % metrics
0
10
20
30
40
Version 3.3
Version 3.4
Version 3.5
Version 3.6
Version 3.7
Version 3.8
Version 3.9
Version 3.10
Versions

Pie chart distribution :

Previous chart but in Tasty pie chart…



Which age groups mostly prefer python…?



Thiobe Index

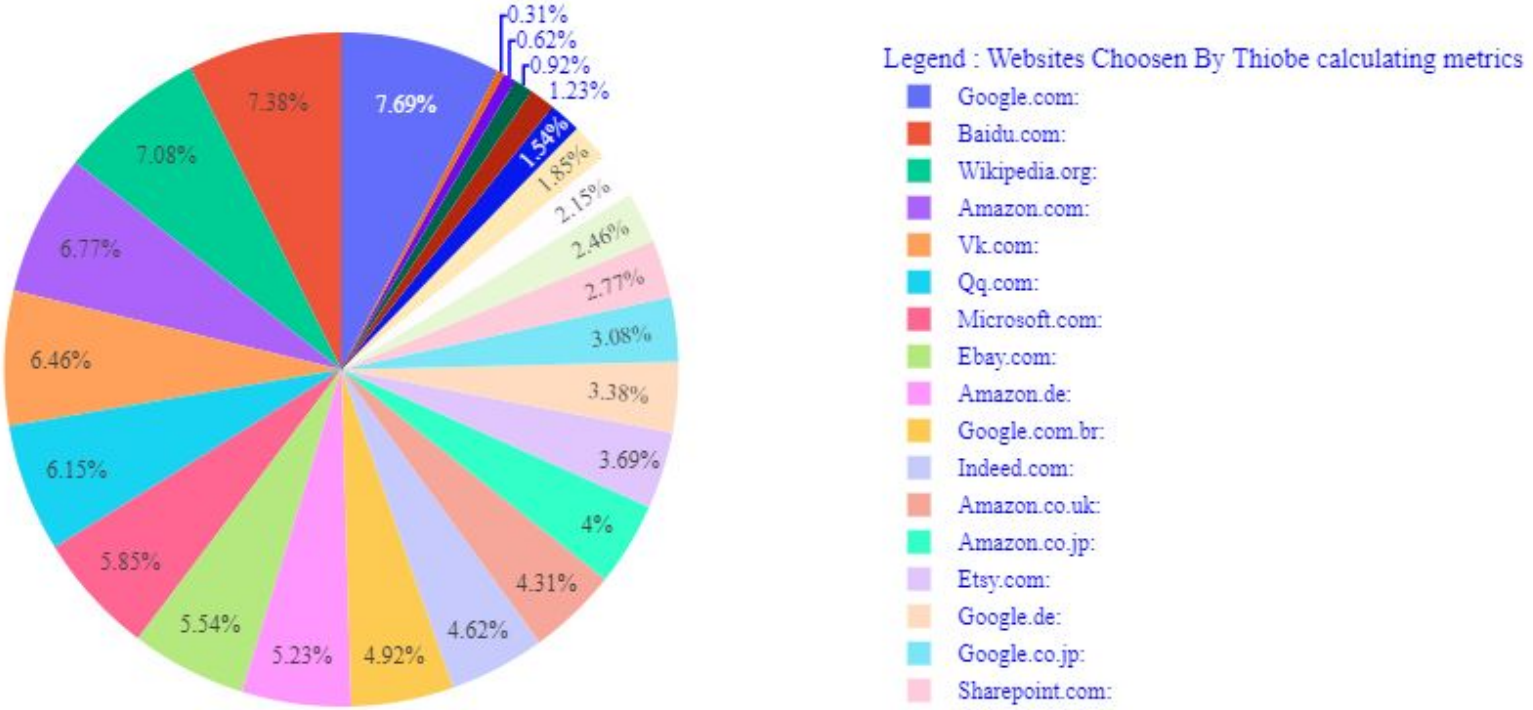
There are 25 search engines that are used to calculate the TIOBE index.

The selected search engines are the 25 highest ranked websites of Alexa that meet the following conditions.

- The entry page of the site contains a search facility.
- The result of querying the site contains an indication of the number of page hits.
- The results should be available in HTML with clear tags.
- Search engines in languages with special characters should be encoded properly.
- The search engine should at least return 1 hit for 1 query.
- The results of querying the site shouldn't contain too many outliers.

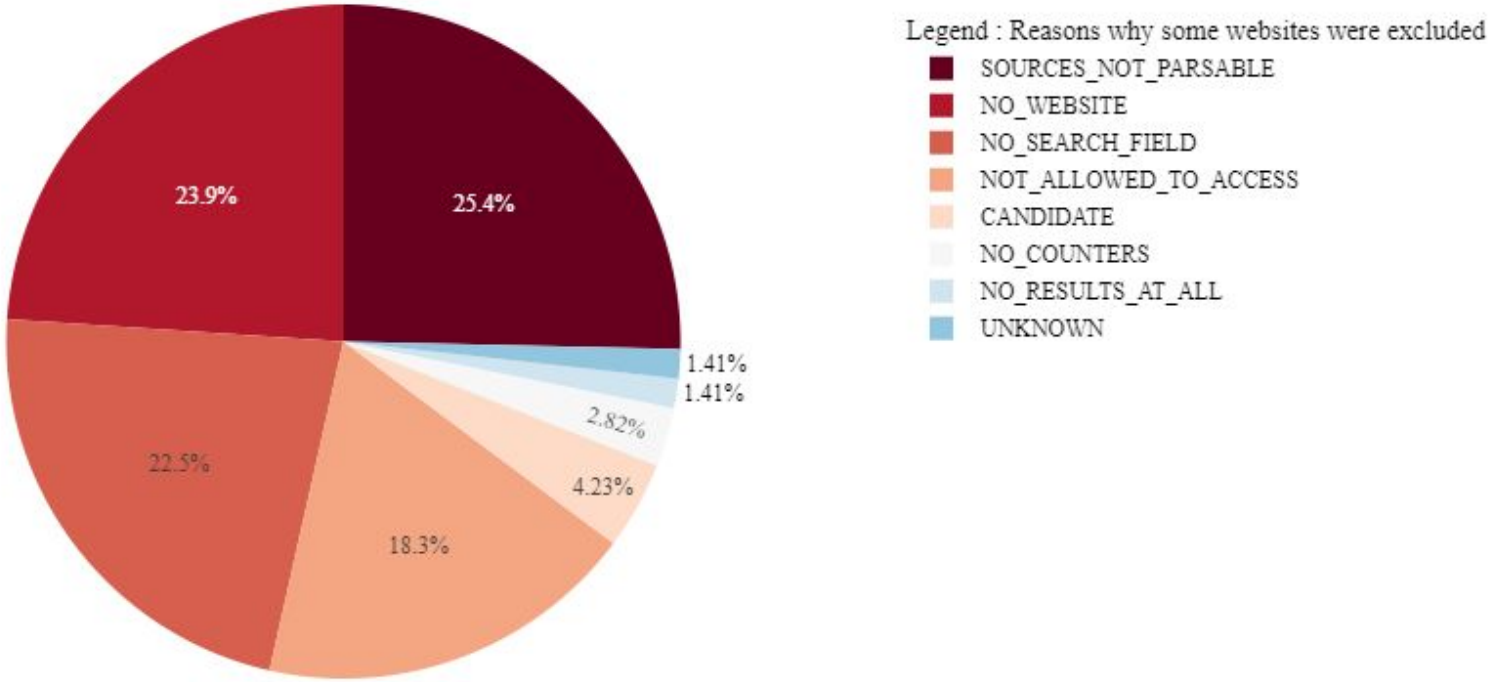
Pie chart distribution :

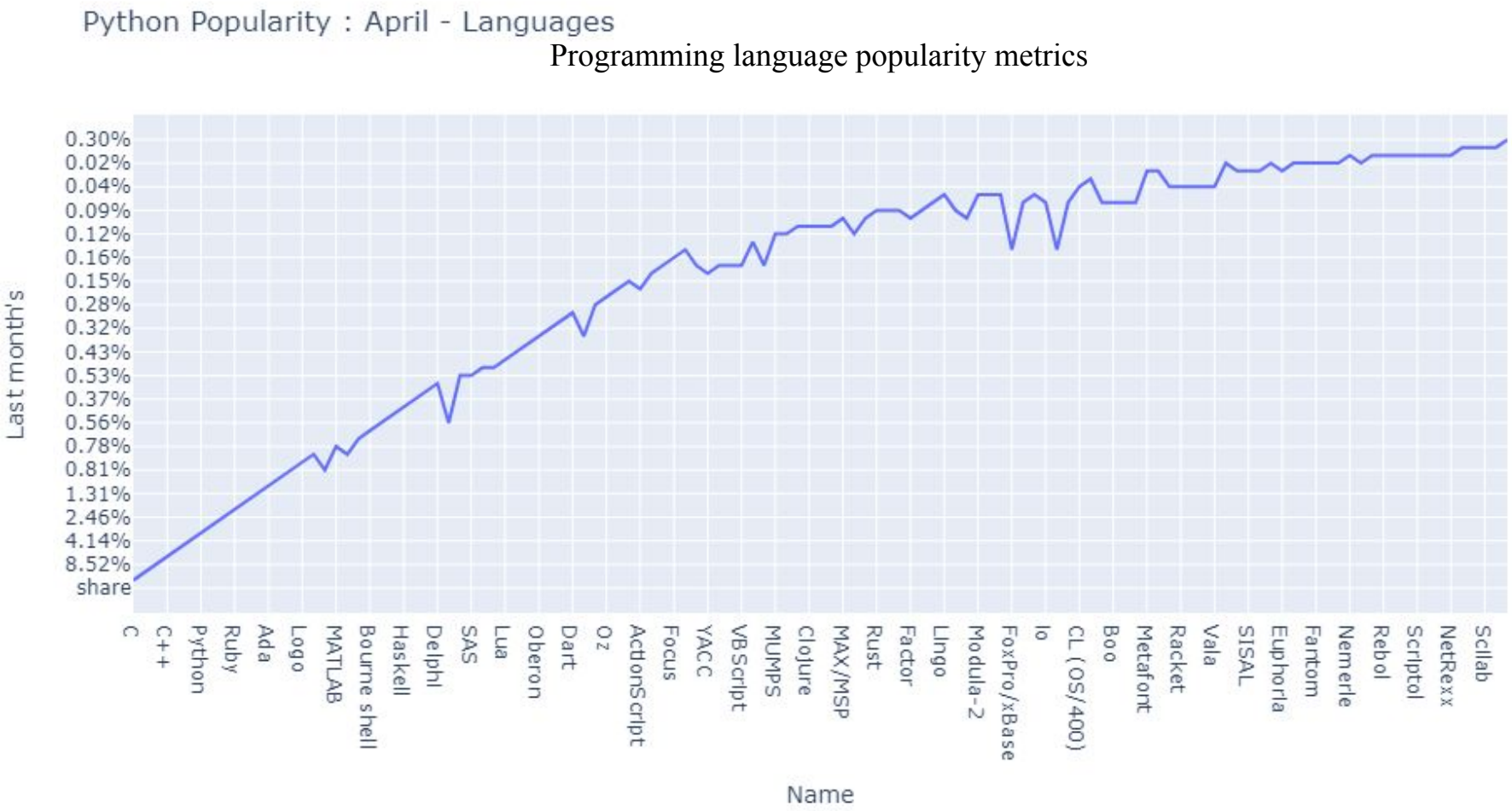
Sites which were approved for data collection by THIOBE index

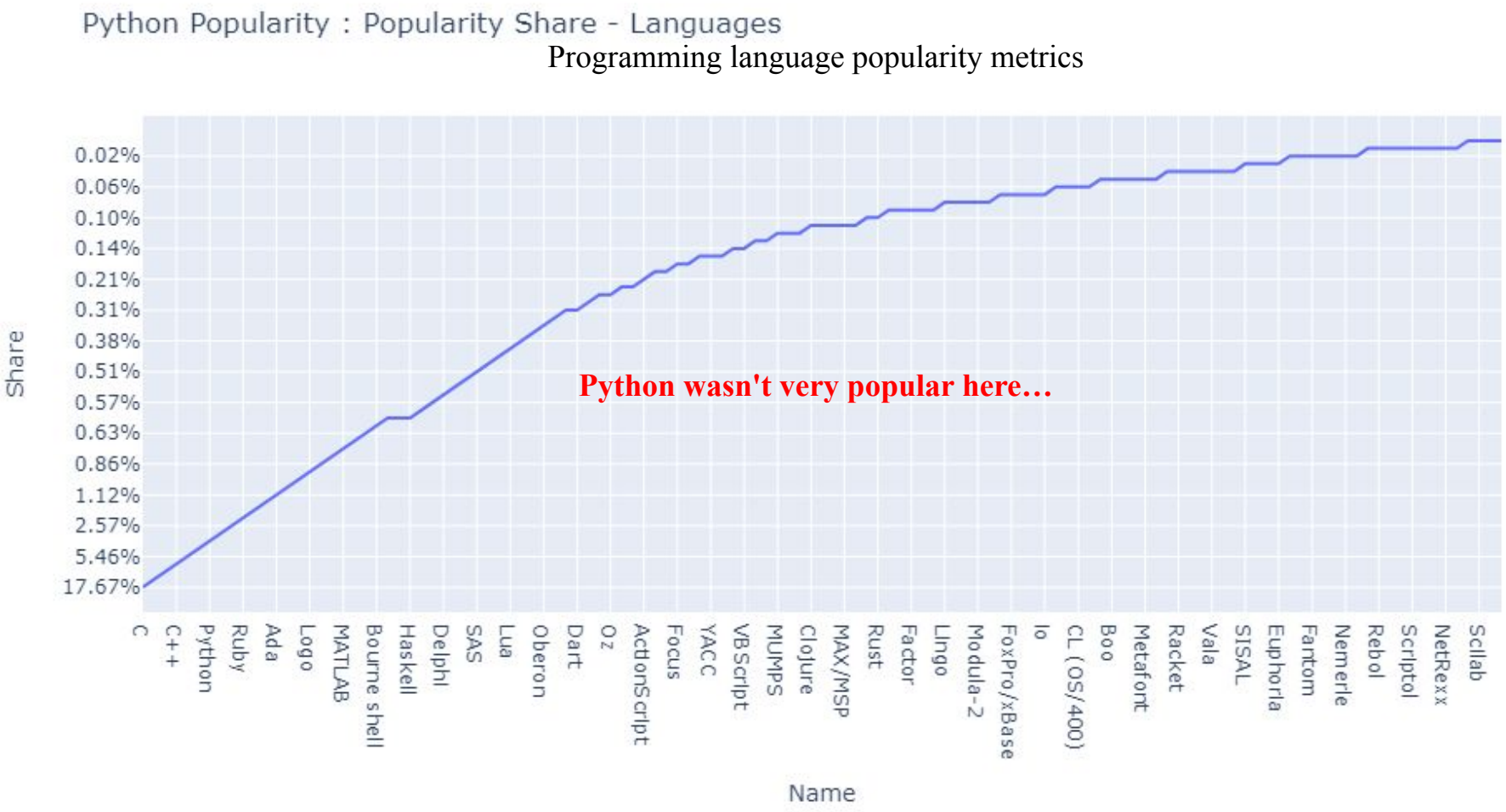


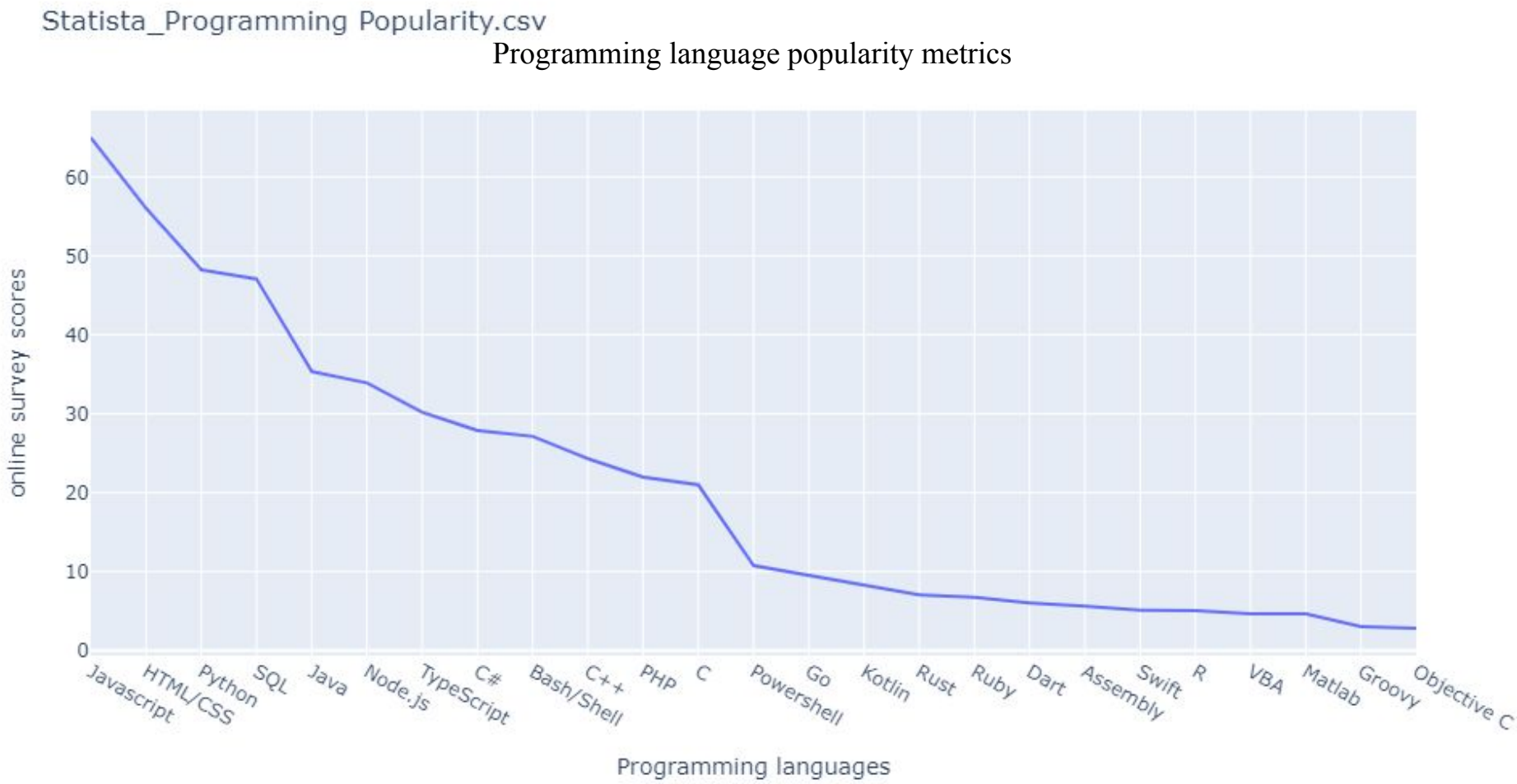
Pie chart distribution :

A large portion didn't make the cut.



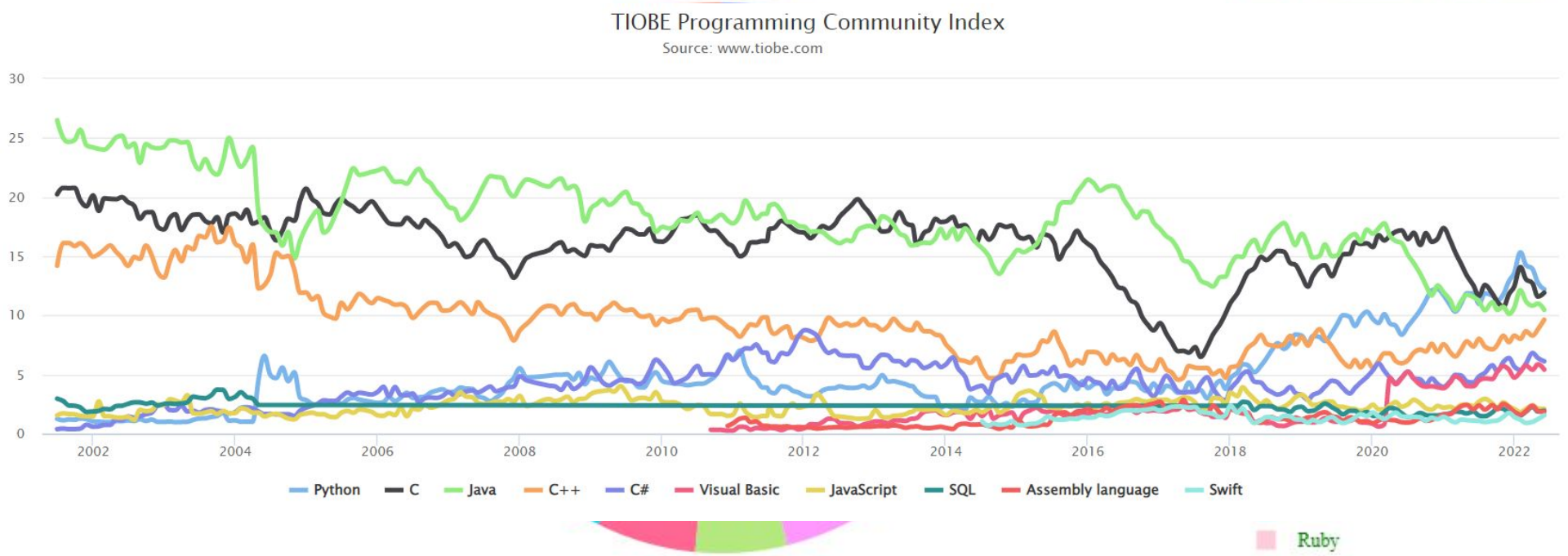

Python Popularity : April - Languages
Programming language popularity metrics
Last month's share
0.30%
0.02%
0.04%
0.09%
0.12%
0.16%
0.15%
0.28%
0.32%
0.43%
0.53%
0.37%
0.56%
0.78%
0.81%
1.31%
2.46%
4.14%
8.52%
C
C++
Python
Ruby
Ada
Logo
MATLAB
Bourne shell
Haskell
Delphi
SAS
Lua
Oberon
Dart
Oz
ActionScript
Focus
YACC
VBScript
MUMPS
Clojure
MAX/MSP
Rust
Factor
Lingo
Modula-2
FoxPro/xBase
Io
CL (OS/400)
Boo
Metafont
Racket
Vala
SISAL
Euphoria
Fantom
Nemerle
Rebol
Scriptol
NetRexx
Scilab
Name


Python Popularity : Popularity Share - Languages
Programming language popularity metrics
Share
0.02%
0.06%
0.10%
0.14%
0.21%
0.31%
0.38%
0.51%
0.57%
0.63%
0.86%
1.12%
2.57%
5.46%
17.67%
Python wasn't very popular here…
C
C++
Python
Ruby
Ada
Logo
MATLAB
Bourne shell
Haskell
Delphi
SAS
Lua
Oberon
Dart
Oz
ActionScript
Focus
YACC
VBScript
MUMPS
Clojure
MAX/MSP
Rust
Factor
Lingo
Modula-2
FoxPro/xBase
Io
CL (OS/400)
Boo
Metafont
Racket
Vala
SISAL
Euphoria
Fantom
Nemerle
Rebol
Scriptol
NetRexx
Scilab
Name

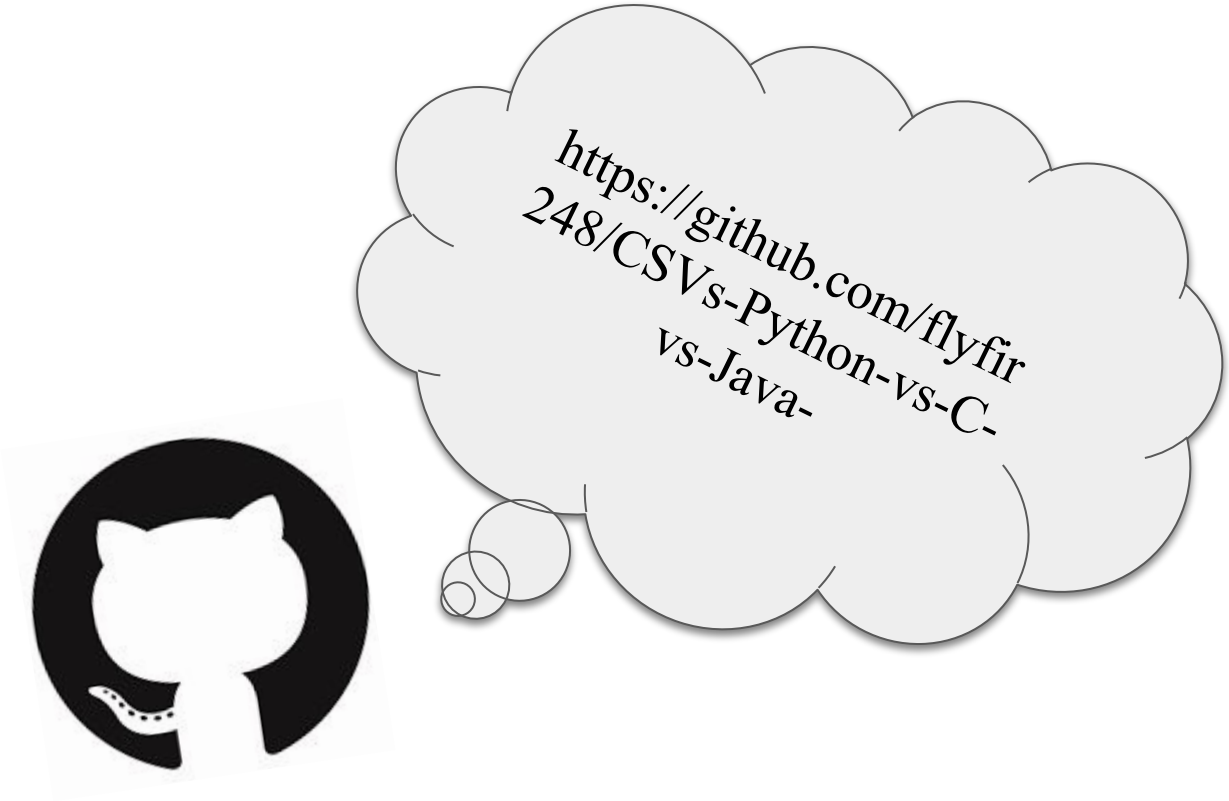

Statista_Programming Popularity.csv
Programming language popularity metrics
online survey scores
0
10
20
30
40
50
60
Javascript
HTML/CSS
Python
SQL
Java
Node.js
TypeScript
C#
Bash/Shell
C++
PHP
C
Powershell
Go
Kotlin
Rust
Ruby
Dart
Assembly
Swift
R
VBA
Matlab
Groovy
Objective C
Programming languages

Pie chart distribution :

Again… a tasty pie chart for you…

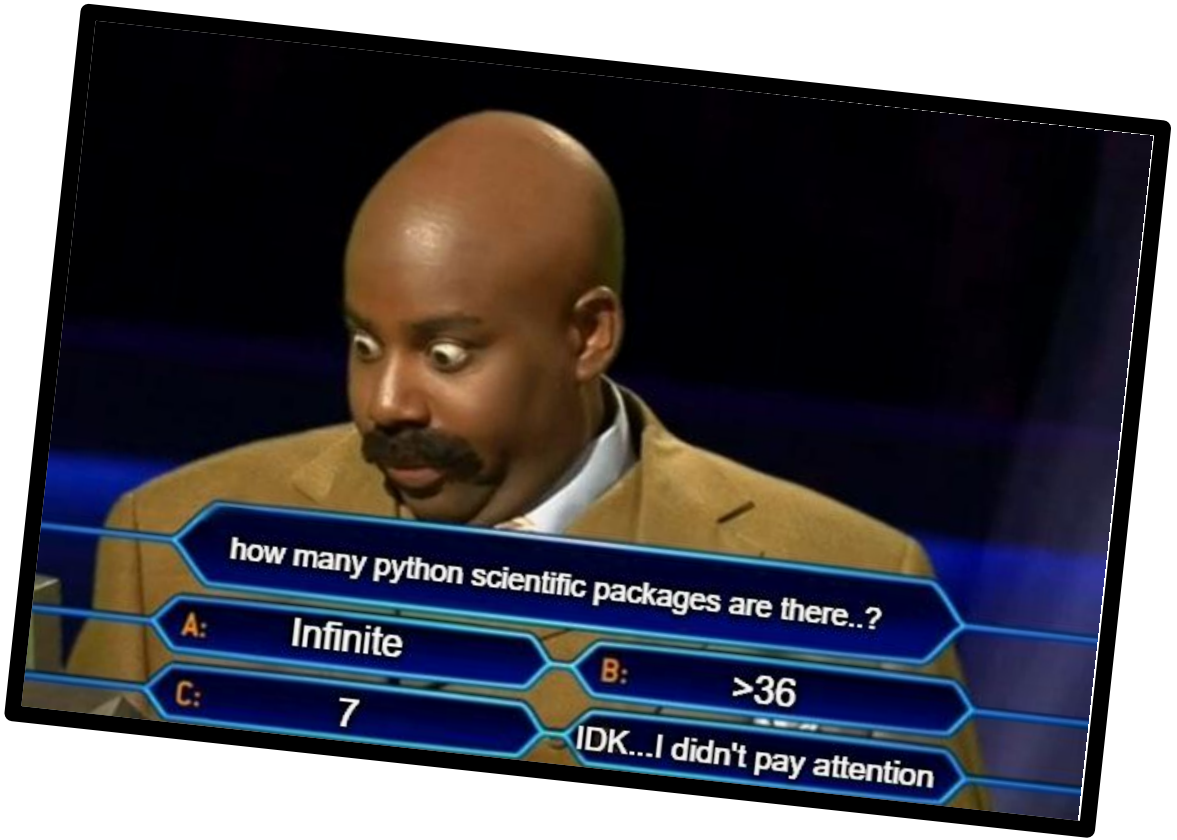


Access these resources here..!!

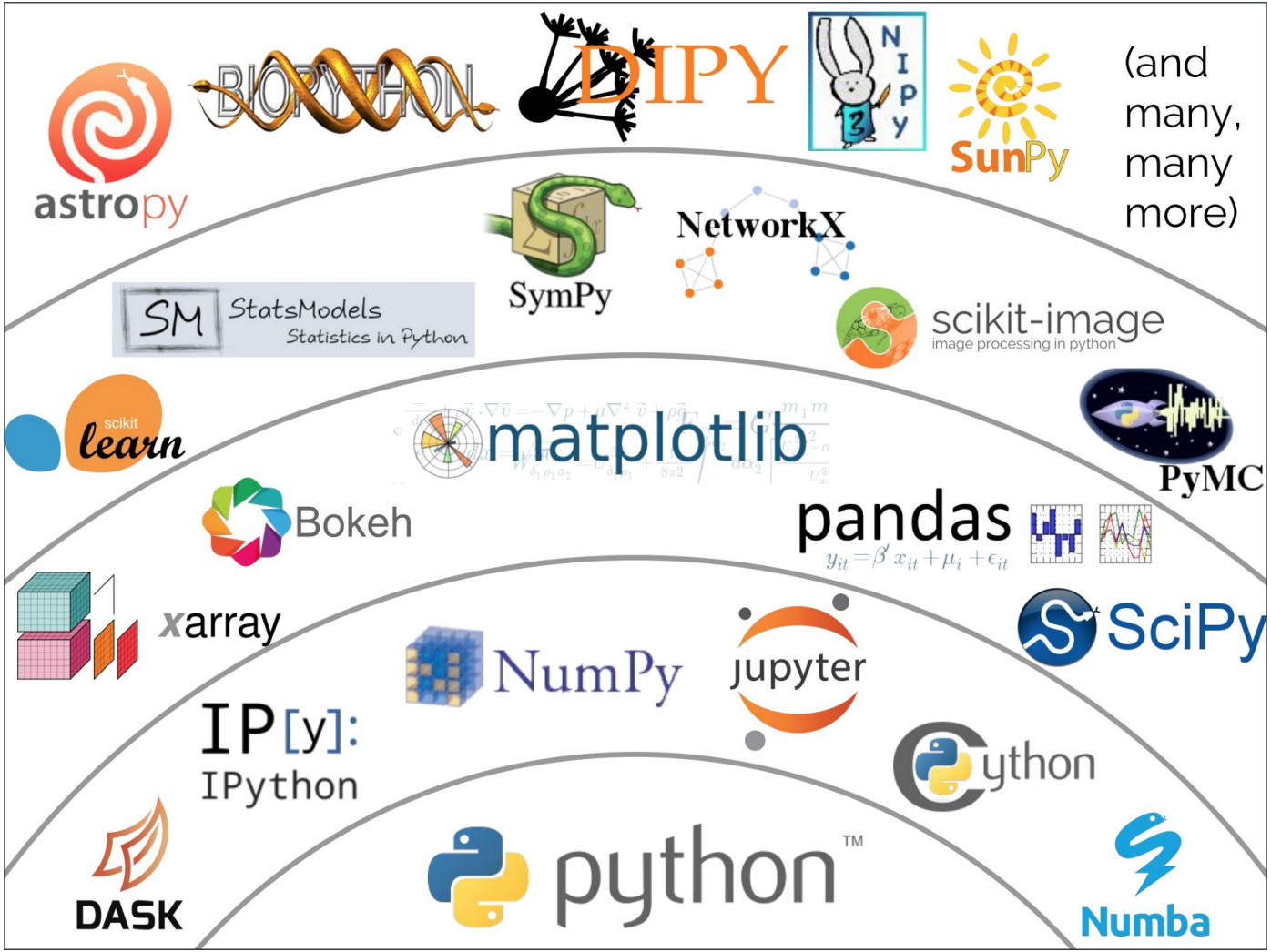


Content

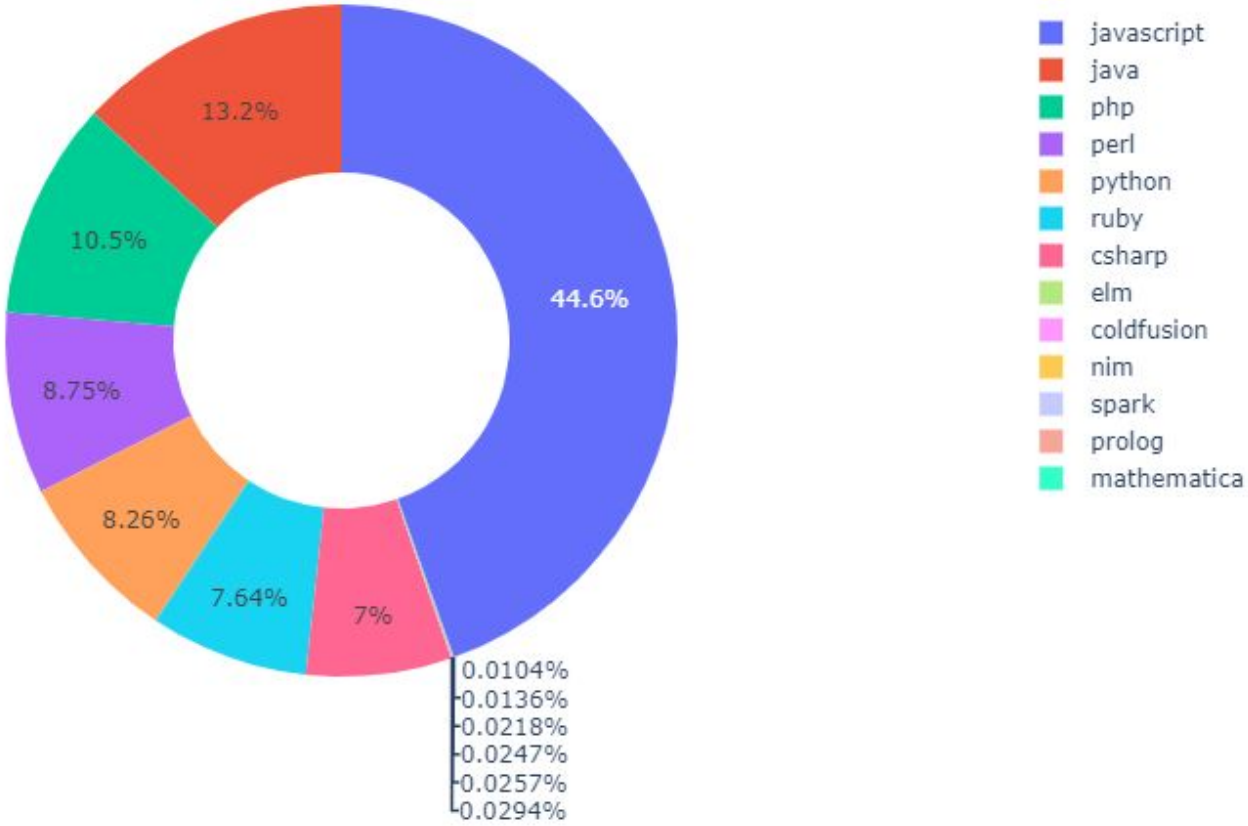
Python Packages !!!




Java


astropy
BIOPYTHON
DIPY
NIPY
SunPy
(and many, many more)
SymPy
NetworkX
SM StatsModels Statistics in Python
scikit-image
image processing in python
scikit learn
matplotlib
PyMC
Bokeh
pandas
xarray
SciPy
NumPy
jupyter
IP[y]: IPython
ython
DASK
python™
Numba

The same as before but as a skinny donut



Some of the Packages

Astropy is a collection of astronomy-related software tools.

Biopython is an open-source Python library for computational biology and bioinformatics.

Bokeh is a Python interactive visualization package that displays content in modern web browsers. (**Cubes**)

Dask is a parallel computing package for analytic computing.(**DEAP/DataMelt**)

Matplotlib is a Python 2D charting package that generates high-quality figures

Mlpy is a machine learning library based on the GNU Scientific Libraries NumPy/SciPy.

NumPy is a Python module that adds support for massive, multidimensional arrays and matrices. (**Pandas**)

Pipenv - Virtual environments.

Some of the Packages

PsychoPy is a Python tool for creating neuroscience and experimental psychology investigations.

PySpark is Apache Spark Python API.

Weka is a Java-based machine learning software suite developed by New Zealand's University of Waikato.

PyTorch is a deep learning framework that allows you to experiment quickly and easily.

ScientificPython is a collection of scientific computing modules written in Python.

Scikit-image is a library for image processing.

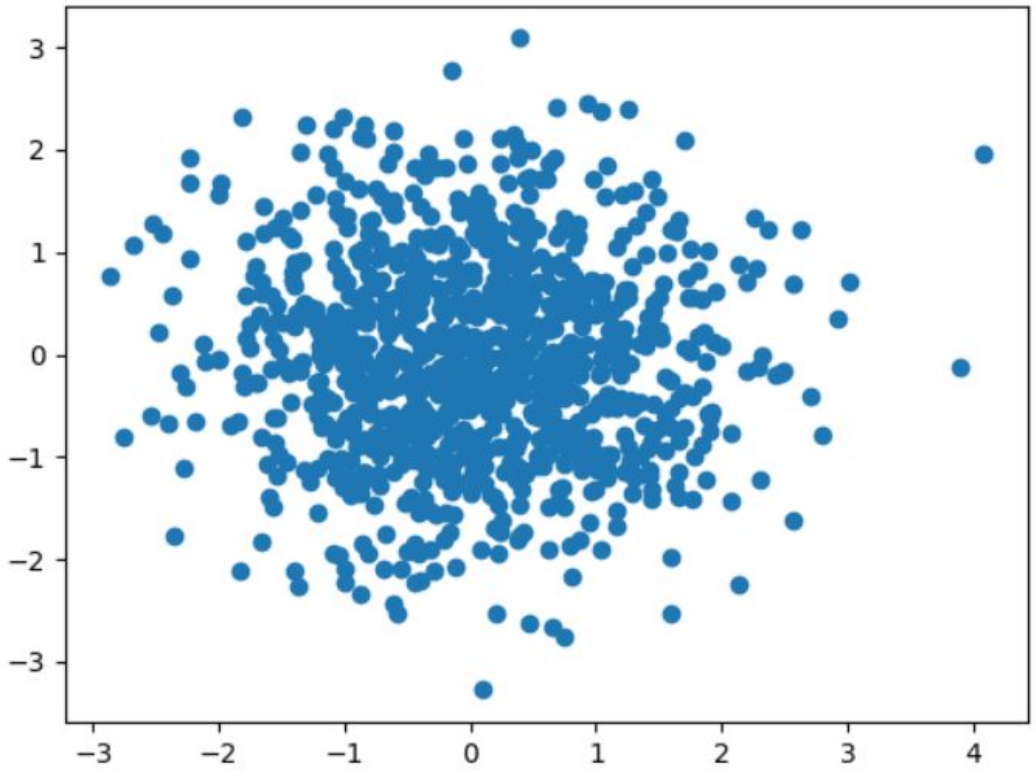
Scikit-learn is a library for machine learning.

SciPy is a scientific and technical computing library used by scientists, analysts, and engineers.

Seaborn is a popular Python data visualization toolkit for creating statistical visuals.

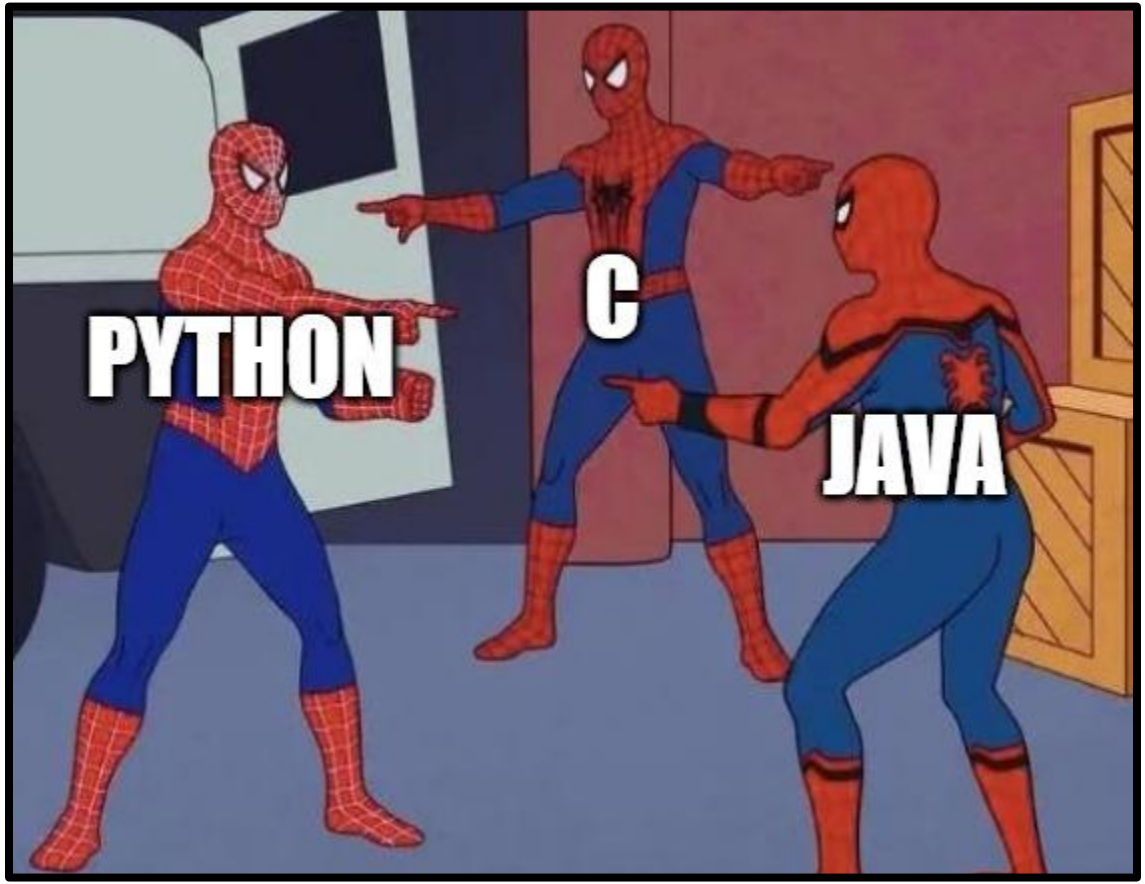
File | C:/Users/anoop/SEM%202/z_PPT/py.html

Let's plot random numbers



Content

Python V Java V C



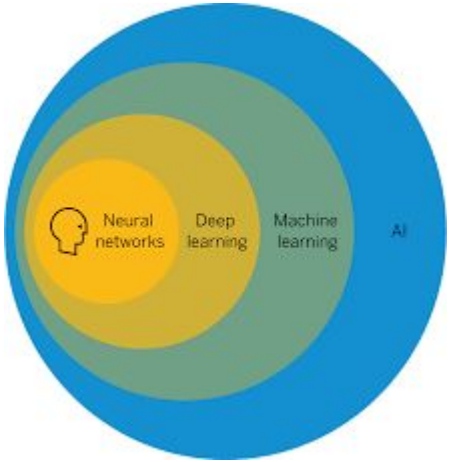
Python V Java V C

Rank	Language	Type	Score
1	Python		100.0
2	Java		95.4
3	C		94.7
4	C++		92.4
5	JavaScript		88.1
6	C#		82.4
7	R		81.7
8	Go		77.7
9	HTML		75.4
10	Swift		70.4

So is python the best programming language.. ever…??

No… it depends on where you are using these languages

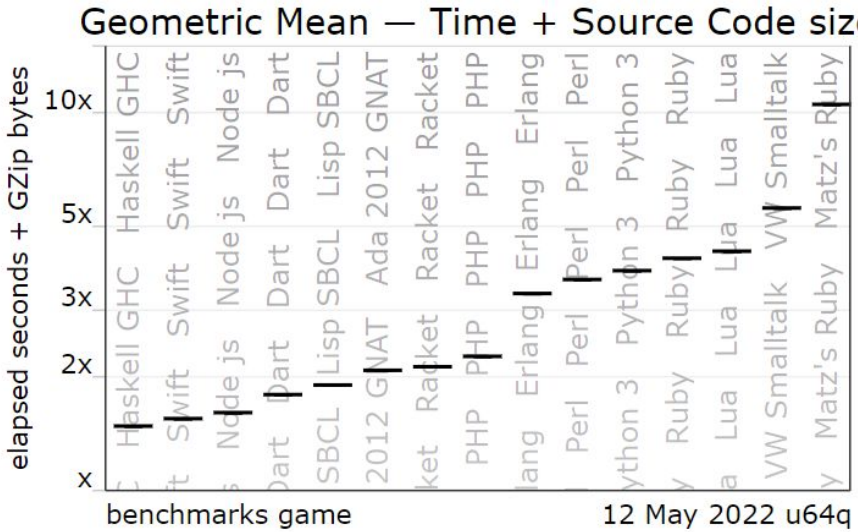
Python V Java V C



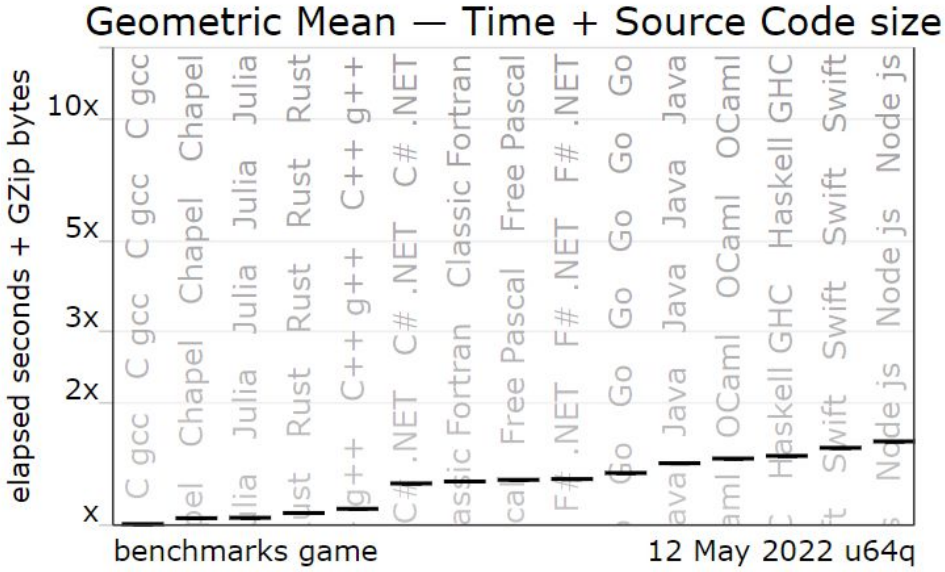
UNREAL
ENGINE

Content

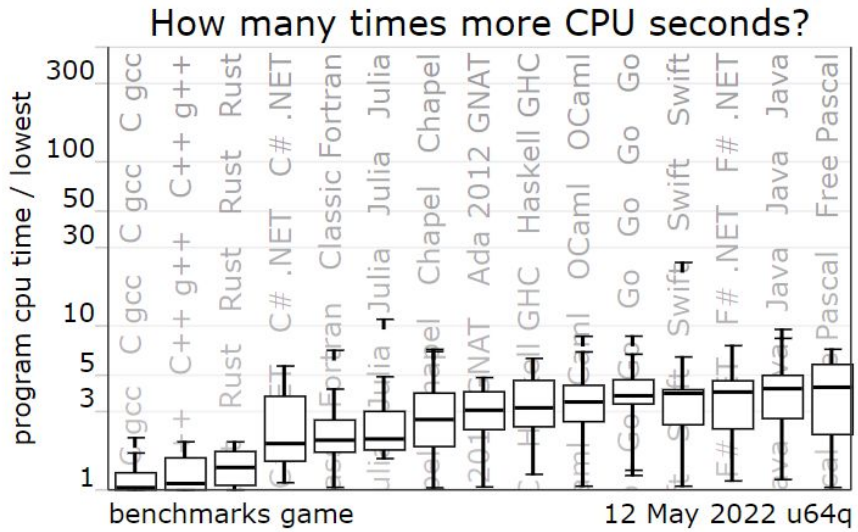
marking

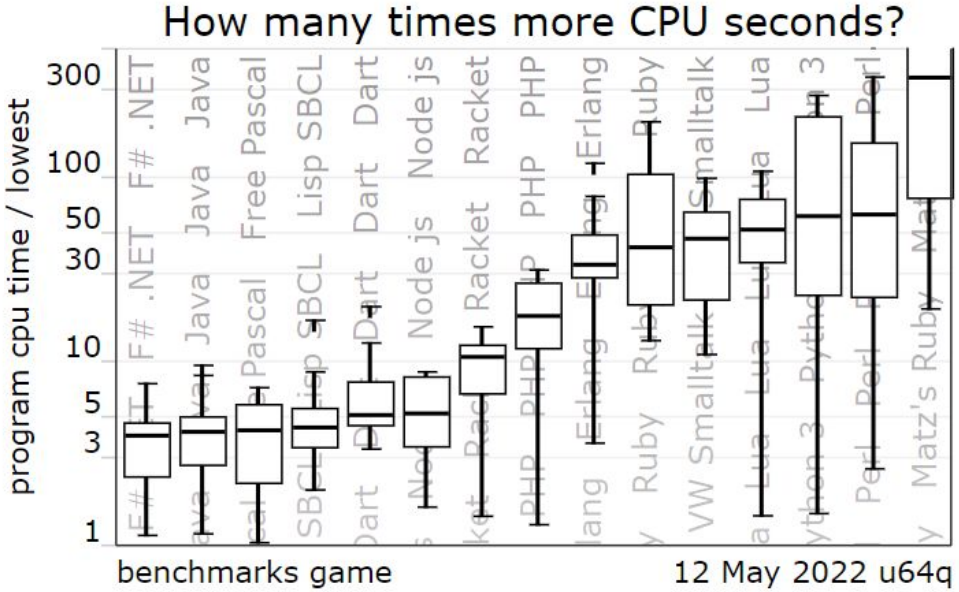

Geometric Mean — Time + Source Code size
elapsed seconds + GZip bytes
10x
5x
3x
2x
x
Haskell GHC
Swift
Node js
Dart
Lisp SBCL
2012 GNAT
Racket
PHP
Erlang
Perl
Python 3
Ruby
Lua
VW Smalltalk
Matz's Ruby
benchmarks game
12 May 2022 u64q


80
100


Geometric Mean — Time + Source Code size
elapsed seconds + GZip bytes
10x
5x
3x
2x
x
C gcc
Chapel
Julia
Rust
C++ g++
C# .NET
Classic Fortran
Free Pascal
F# .NET
Go
Java
OCaml
Haskell GHC
Swift
Node js
benchmarks game
12 May 2022 u64q

marking

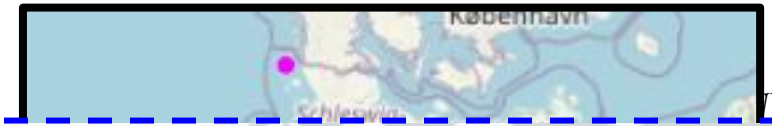

How many times more CPU seconds?
program cpu time / lowest
300
100
50
30
10
5
3
1
C gcc
C++ g++
Rust
C# .NET
Classic Fortran
Julia
Chapel
Ada 2012 GNAT
Haskell GHC
OCaml
Go
Swift
F# .NET
Java
Free Pascal
benchmarks game
12 May 2022 u64q

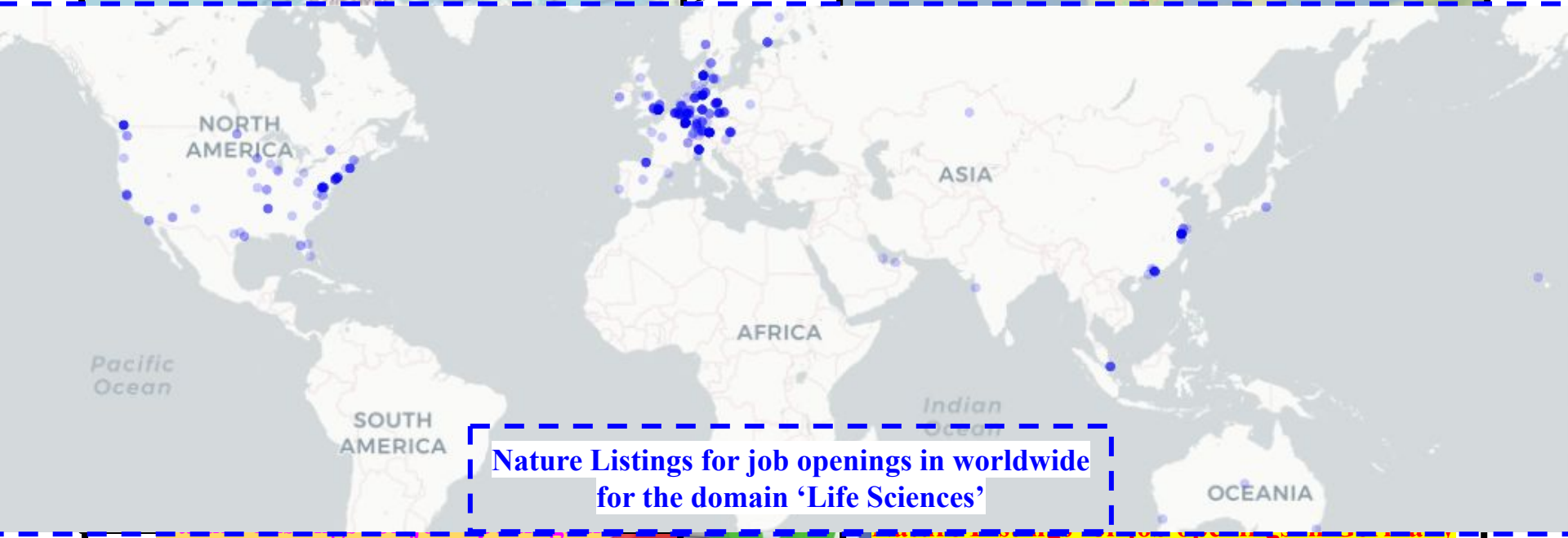

How many times more CPU seconds?
program cpu time / lowest
300
100
50
30
10
5
3
1
F# .NET
Java
Free Pascal
Lisp SBCL
Dart
Node js
Racket
PHP
Erlang
Ruby
Smalltalk
Lua
Python 3
Perl
Matz's Ruby
benchmarks game
12 May 2022 u64q

Content

Conclusion

What about Jo





Nature Listings for job openings in worldwide for the domain 'Life Sciences'

Germany for the domain 'Bioinformatics'

for the domain 'Life Sciences'

Conclusion

Popularity of python is currently on an all time high.

Most users find python more easier to use.

Widely available packages and community support makes it more usable.

We probably will be seeing more python in the near future.

"There are only two kinds of programming languages: those people always bitch about and those nobody uses."
Bjarne Stroustrup.

Content

References

Python Versions | Python.org. (n.d.). Retrieved May 13, 2022, from https://www.python.org/doc/versions/

The Most Popular Python Scientific Libraries. (n.d.). Retrieved May 13, 2022, from https://www.stxnext.com/blog/most-popular-python-scientific-libraries/

Will Python 2 Ever Die? | Python in 1 minute. (n.d.). Retrieved May 13, 2022, from https://pythonin1minute.com/will-python-2-ever-end/

23+ Awesome Python Statistics 2021. (n.d.). Retrieved May 14, 2022, from https://leftronic.com/blog/python-statistics/

Usage Statistics and Market Share of Python Version 3 for Websites, May 2022. (n.d.). Retrieved May 13, 2022, from https://w3techs.com/technologies/details/pl-python/8

Most used languages among software developers globally 2021 | Statista. (n.d.). Retrieved May 14, 2022, from https://www.statista.com/statistics/793628/worldwide-developer-survey-most-used-languages/

The Transparent Language Popularity Index. (n.d.). Retrieved May 13, 2022, from http://lang-index.sourceforge.net/

Programming Languages Definition | TIOBE - The Software Quality Company. (n.d.). Retrieved May 13, 2022, from https://www.tiobe.com/tiobe-index//programming-languages-definition/#instances

Mertz, David. (n.d.). *Picking a Python version : a manifesto : from_future_import Python*.

Maciej Serda. (2013). Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza. *Uniwersytet Śląski*, 343–354. https://doi.org/10.2/JQUERY.MIN.JS

Does every programming language have a central package repository? | CodeLani. (n.d.). Retrieved May 16, 2022, from https://codelani.com/posts/does-every-programming-language-have-a-central-package-repository.html

We thank and appreciate you for your attention ..!!

Clap Please..

Scientific Presentation :

Outlook on the development of Python

Presenter : Michael Lee

Semester 2

Masters Student
Life Science Informatics
Life Science Informatics Department
University of Bonn

Questions I will try to answer today:

1. What is the outlook on the development of Python?
2. Are there other interesting languages for the LSI Community?
3. If so, which are they and why are they well suited for LSI research?
4. What is the preferred programming language in LSI?
5. Is it one programming language or is it a combination of programming languages?
6. *What are other programming tools (other than languages) that are required by employers when working in LSI?

Content

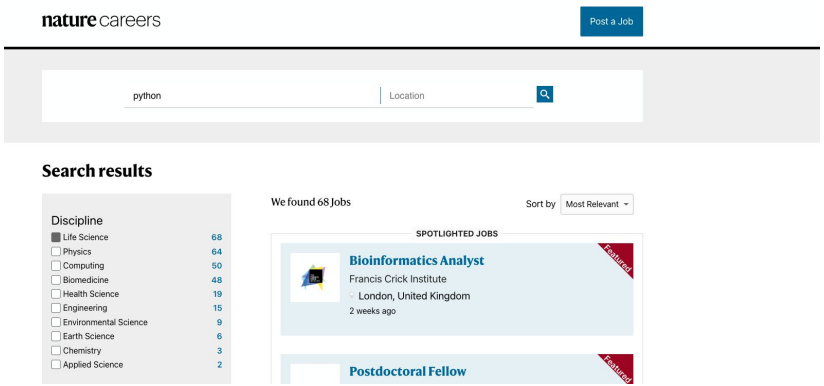
Outlook on the development of Python

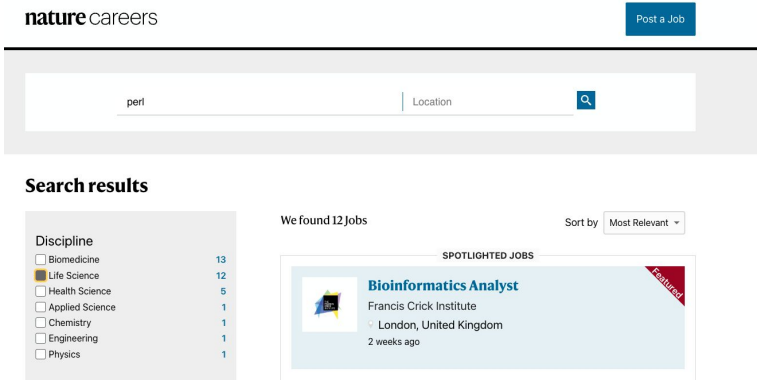
Docs by version

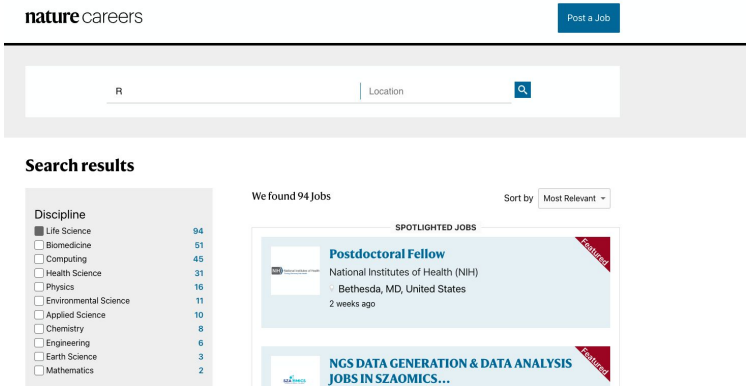
Python 3.12 (in development)
Python 3.11 (pre-release)
Python 3.10 (stable)
Python 3.9 (stable)
Python 3.8 (security-fixes)
Python 3.7 (security-fixes)
Python 3.6 (EOL)
Python 3.5 (EOL)
Python 2.7 (EOL)
All versions

- Current status as of 16.06.2022 from the official documentation at docs.python.org.
- Python creator Guido van Rossum has speculated that there will never be a Python version 4 released.

Searching Nature Journal's career page







Other interesting languages in the LSI community

“Ideal candidate must have strong linux shell coding experience, comfortable in using R to do data analysis, and proficient in Python.” - job requirement results of keyword search “bioinformatics” into glassdoor.com.

The most commonly used languages in LSI

Python:

- Scalability advantage when working with large datasets (BAM files upwards of 100’s of GB’s). Python packages have been shown to keep RAM usage more efficient (ex. Scanpy).
- Good interactive script support (Jupyter Notebook)
- Advantage in converting to API endpoints (ex. FLASK, Django REST AP)

R:

- More options for plotting packages built specifically for genomics data (ex. ggplot2)
- More robust ecosystem of pre existing packages *for bioinformatics* (ex. Seurat).
- Also good interactive script support (Markdown files - similar to Jupyter Notebooks)

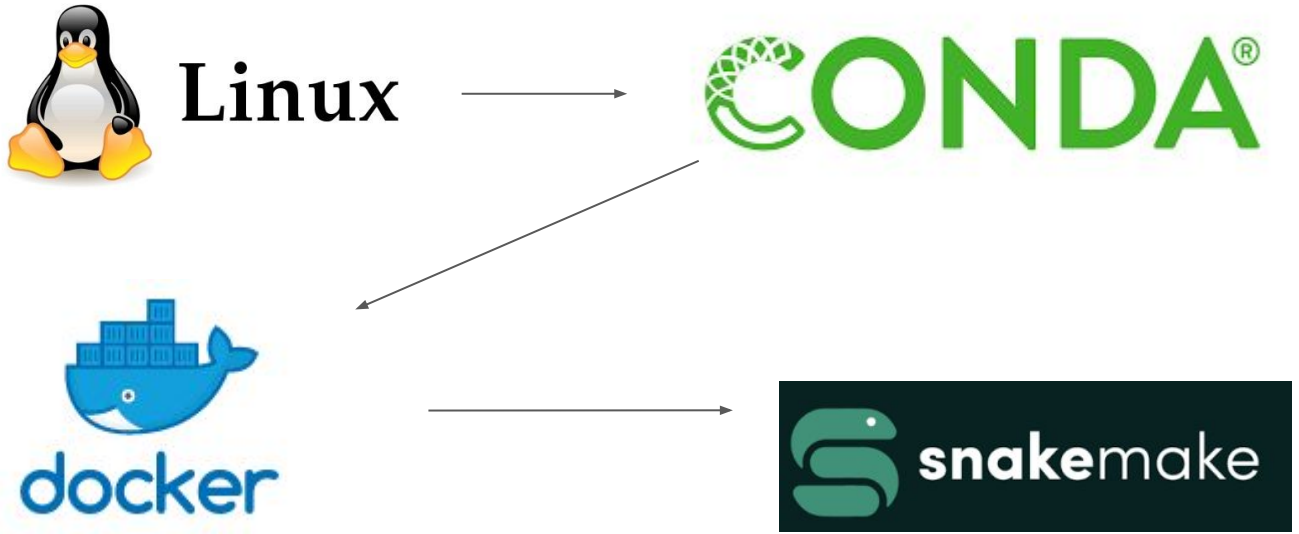
Perl, another language of interest

- A high-level general purpose, interpreted, programming language.
- Syntactically similar to C.
- Perl code has been reported to be more difficult to read than python code.
- Python is much more scalable and therefore more popular in the LSI community.

Preferred language in LSI

- It depends.
- In general…
  - For scalability and processing large datasets, it is common to use Python.
  - For powerful plotting and visualization of genomics data, many bioinformaticians prefer R.
- But these are just general guidelines, each project in LSI has different needs and depending on what a scientist is trying to achieve Python or R would be a better choice (But that is a topic for a whole different presentation, and actually a topic of many highly cited published papers in LSI).

Other common requirements (software tools)



Linux

- For many bioinformatics projects, a simple laptop will not have enough computing power.
- If you are given access to a high performance computing cluster (HPC), this server, will commonly use Linux as its operating system.
- Knowing the basics of how to navigate and use Linux is a big plus for working in LSI and in programming in general.

- A lot of projects in LSI require different versions of different packages that have a lot of dependencies on other package/software versions.
- For example, different python packages may depend on specific versions of python.
- It order to isolate a project to a specific environment, Conda allows users to build virtual environments specifying which package dependencies are needed.
- Instead of using 1 large growing environment for all projects, create many smaller virtual environments specific to your projects to try keep the package “dependency train” to a minimum
- It is definitely a plus to understand how conda works.

- Docker is a containerization platform that allows you to containerize an entire environment used to run an application, and run the application on any machine with docker installed.
- This is very useful when trying to reproduce results of another project on your machine.
- For scientific publications in LSI, it is necessary to provide a data and software reproducibility section and a docker container is a very common medium for this purpose.

- “Snakemake is a workflow engine that provides a readable Python-based workflow definition language and a powerful execution environment that scales from single-core workstations to compute clusters without modifying the workflow.”
- Coordinates the execution of multiple scripts within a workflow while allowing customizability of environment parameters
- Integrates very well with Conda Package Manager and Docker’s container virtualization to make workflows portable.
- Quick Overview Video

> Bioinformatics. 2012 Oct 1;28(19):2520-2. doi: 10.1093/bioinformatics/bts480. Epub 2012 Aug 20.

Snakemake--a scalable bioinformatics workflow engine

Johannes Köster [1], Sven Rahmann

Affiliations + expand

PMID: 22908215 DOI: 10.1093/bioinformatics/bts480

Summary

- Python3 will be around for many years
- Python and R are the top programming languages to use in the LSI community
- As seen in job advertisements, other supporting tools are required
- Many employers seek applicants with Linux experience
- Use Conda to manage different environments which may require different package versions
- Use Docker to make environments for data analyses reproducible
- Use Snakemake to orchestrate workflows

References

https://docs.python.org/3/

https://www.nature.com/naturecareers

https://www.redhat.com/en/topics/linux/what-is-linux#:~:text=Linux%C2%AE%20is%20an%20open,resources%20that%20do%20the%20work.

https://docs.conda.io/en/latest/

https://docs.docker.com/

https://snakemake.readthedocs.io/en/v5.1.4/index.html