

What is the Best Programming Language to Learn in 2024?

Ruhaan Shinde, Purdue University, December 2024

There's a lot of speculation over which programming language is actually the most useful to learn in 2024. Is it Python? Java? Or is SQL preferred with the insane volume of datasets nowadays? To answer this question, I have looked at datasets surrounding programmer's preferences, and how those preferences are changing with the modern age.

This dataset is accessed through Kaggle, and will be linked through a Github repository with RStudio code (for data modeling purposes) at the bottom of this paper.

Firstly, this isn't going to be an ordinary analysis of which programming language is the best to learn. It's based on datasets with user preferences, to see which languages are more popular and growing in this day and age. Statistical tests with data that actually backs up these statements.

For my fellow data scientists, stick around if you're interested in knowing how I preprocessed the data using Python and RStudio. Otherwise, keep scrolling down.

Data Preprocessing:

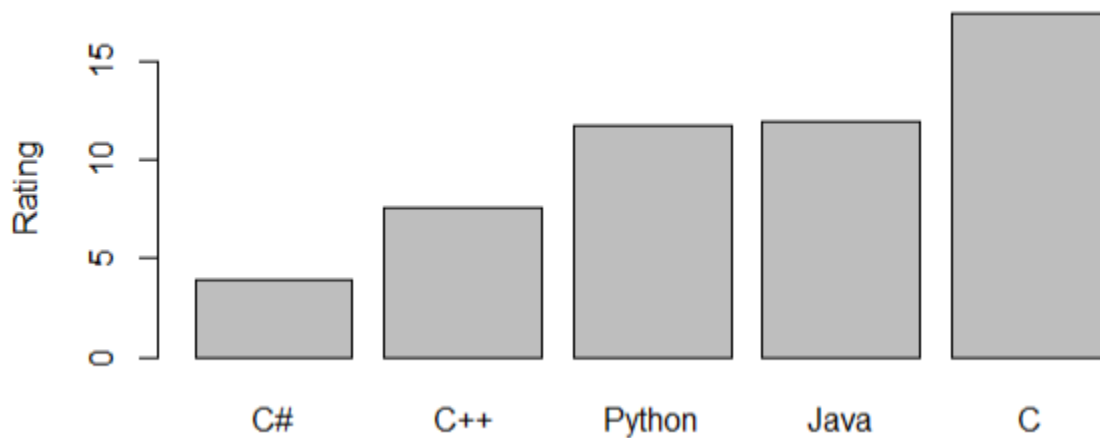
Utilizing Python and RStudio, I filtered through an initial dataset with 20 programming languages and how users rate them. I made 3 smaller datasets, one with results for only the top 5 programming languages, one that highlights the annual change in preference, and one that only highlights user ratings. Given the relatively small volume of data, compiling it into a matrix for

singular value decomposition wasn't necessary. Instead, filtering through dataframes proved sufficient.

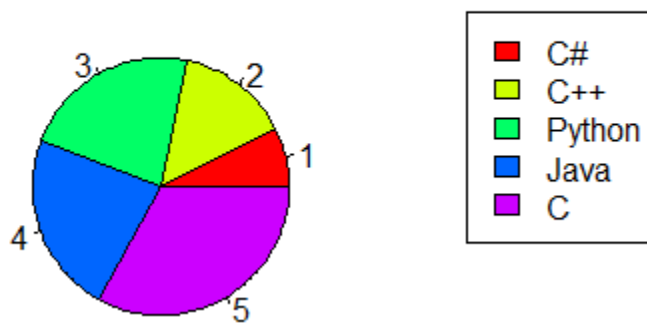
Anyways, the most important step to understand data is visualizing it. This makes it easier to actually know what's going on in these massive sets of tuples and attributes aka tables.

So to answer the question, based on user preferences (from around the world), the most popular language isn't actually Python, Java, or SQL. It's the dreaded C.

5 Languages with Highest Ratings



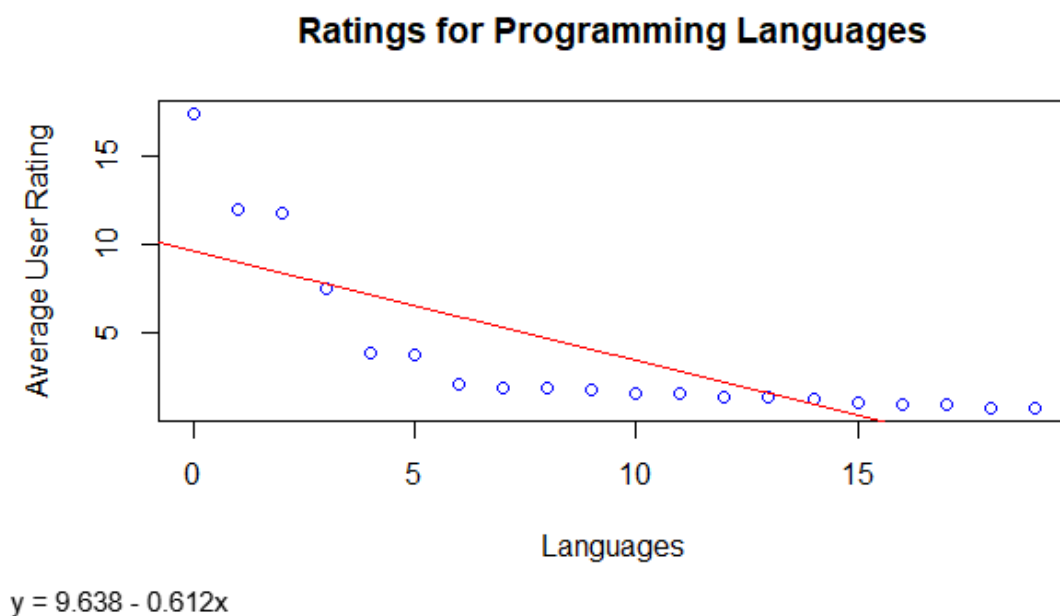
Top 5 Popular Languages



*Numbers refer to Language ID in the dataframe, not the ranking of popularity

A basic R barplot and pie plot above show that C has significantly higher user ratings, upwards of 15.0 (ratings are based on a scale of 20.0). Java and Python were rated pretty similarly, with C# and C++ coming next. Basically, traditional SWE still rules, even with the rise of AI and machine learning techniques. Another sign to practice Leetcode everyday. The basics of computer science will always be fundamental.

Just to visualize how important these 5 languages are in relation to the 15 others, take a look at this comparison of the top 5's ratings against the rest of the languages in the dataset:



A scatterplot of the ratings data show that C, Java, Python, and C++ are significantly preferred over others. A curved slope would show diminishing returns for every single programming language rated below C++, including C#. However, a factor that needs to be taken into account in this study is the age range of users who voted. Senior engineers and any older voters would probably be more likely to prefer such languages, since they are older and more streamlined in

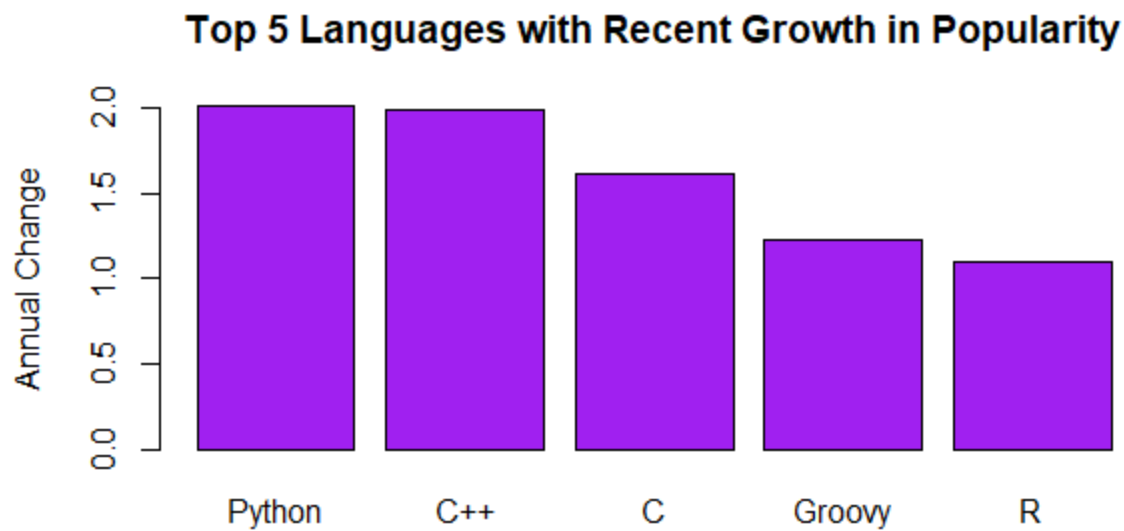
general. Therefore, to truly figure out the most important language in 2024 (because I know you're not convinced that it's C), let's look at the growth rate in preferences of the programming languages in the dataset.

For those of you who like tabular data, here's a ranking list I compiled through dataframe filtering that shows the annual change in ratings of the 20 programming languages studied, after 2022.

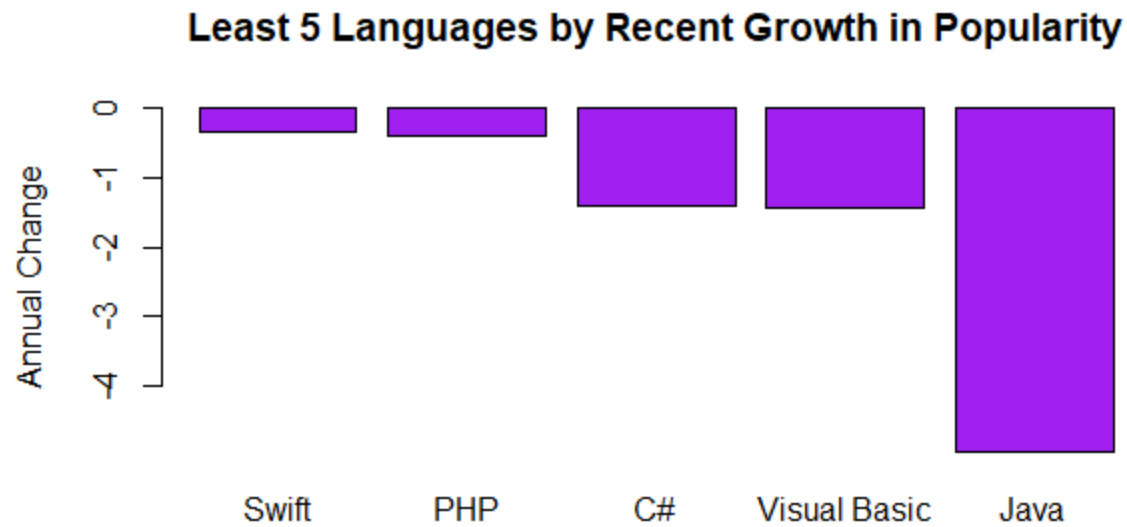
	Programming.Language	Annual.Change
1	Python	2.01
2	C++	1.99
3	C	1.61
4	Groovy	1.23
5	R	1.10
6	Assembly language	0.76
7	Go	0.51
8	MATLAB	0.41
9	Perl	0.27
10	Ruby	0.24
11	SQL	0.10
12	Objective-C	0.07
13	classic visual Basic	-0.04
14	Delphi/Object Pascal	-0.20
15	JavaScript	-0.25
16	Swift	-0.36
17	PHP	-0.41
18	C#	-1.40
19	visual Basic	-1.44
20	Java	-4.93

This table shouldn't surprise anyone. Of course Python is growing the fastest. Python is simply too powerful. It's good for database engineering, analysis, object-oriented programming, and has so many updates and libraries. It's also generally a lot easier to learn than C++, Java, C, or C#.

To look at some of the other languages with growing popularity:



Python and C++ significantly lead the charts, showing how a combination of data analysis and object-oriented programming will pave the way for future programmers. That being said, if you're doing anything data related I recommend you learn R. R helped me with data modeling for this paper and it's way easier to learn and write scripts with than Python. Though R is limited in comparison to Python, for the early steps in the data pipelining process I would suggest using R over Python, if possible. The recent growth of Groovy shows a movement from traditional object-oriented programming, since Groovy makes Java more efficient, and Java is declining. While not necessarily right, Groovy may become a tool of the future. Anyone interested in OOP and SWE should look into this tool.



Speaking of Java, wow. Is Java really declining that much? It's still the primary language taught in first-year CS college courses, yet I've noticed that I have never needed to know it in the workplace. Java truly is becoming useless. C and Python can cover all aspects of OOP, including extra packages for other usages (most importantly databases, API integration, and client-server communication). Python especially does just these things more efficiently, and it is better to be fluent in Python around other engineers just like how you'd want to know English regardless of where you travel in the world. I'm not saying don't learn Java, but it shouldn't be your primary language anymore. Learn C or even C++ instead when it comes to OOP.

The conclusion is: it depends on what walk of computer science you're actually studying. But for starters, Python is non-negotiable. It doesn't matter if you're working in SWE, data science, analytics, etc. Python is the English of the computer science world. Learn it. It will serve you well and you will be able to collaborate with many more developers.

Now to dive deeper - if you're working strictly in SWE, maintaining/integrating applications, working heavily with data structures, etc., it will serve you well to know either C, C++, or C#, in that order. Java is useful too, but will become obsolete in the future. The data backs this up too. And do leetcode every day.

If you're like me and prefer the data side of things, Python is undeniably the cornerstone language for our work. BUT, this isn't monogamous. Despite not being highly ranked in the datasets, I would still suggest learning SQL. SQL is not only important for database interviews but it is also the main database tool and having a strong foundational knowledge of it can get you any entry-level or internship role. Do SQL leetcode too. You might be asking, why SQL when Pandas makes it easier to filter through tuples and attributes? The answer is that SQL is just more accepted and so much easier to use. Pandas is honestly kind of useless outside of a college course where you're forced to learn it. Otherwise, having a basic knowledge should suffice, especially when A.I. can help. For visualizations and statistics, R and Python are both non-negotiable tools to learn, I don't think that needs an explanation though. R doesn't involve much learning though, just doing. That's your cue to download RStudio now and start scripting.

If you know a thing or two about statistics, or want to, I performed a global simple linear regression F-test to see if the annual change in ratings after 2022 actually has an effect on the overall ratings for each language. Interestingly, with a p-value of 0.9308, the annual change has almost little to no significance on the overall rating. Therefore, just because Java is slowing down in popularity doesn't mean you don't learn it. You don't have to, but if you do you aren't

wasting your time, you can just be using your time for better things. If you want to see the in-depth summary of the results, view the Github repository linked below.

Sources:

Repository for data and code (view README):

<https://github.com/ruhaans05/Best-Programming-Language-2024>