PROGRAMMING LANGUAGES VISUALISATION

Rosa Pham

Table of Contents

Introduction

Programming languages have been a crucial aspect of modern technology. They have been developed to address various requirements for developers and created software applications that power numerous facets of everyday life [1]. A programming language comprises a set of rules and instructions that can be executed to solve from simple to complicated tasks. Choosing the right programming language can considerably affect the software development procedure and the developer career path. According to the U.S Bureau of Labor Statistics [2], employment opportunities for programming jobs are predicted to grow by 21% from 2018 to 2028.

Hence, the aim of this visualization report is to offer an understanding of inquiries such as which programming languages are on the top of the ranking used by users, what is the trend and future of programming languages, which programming languages developers should learn to get better job opportunities based on the job requirements. The intended audience may include technical managers who decide which languages for new projects, job seekers from the non-related programming area who want to change their career path, and developers who keep up with the new trend or get pursue better career prospects in new positions.

Visualisation Results

The top 10 languages used the most by users

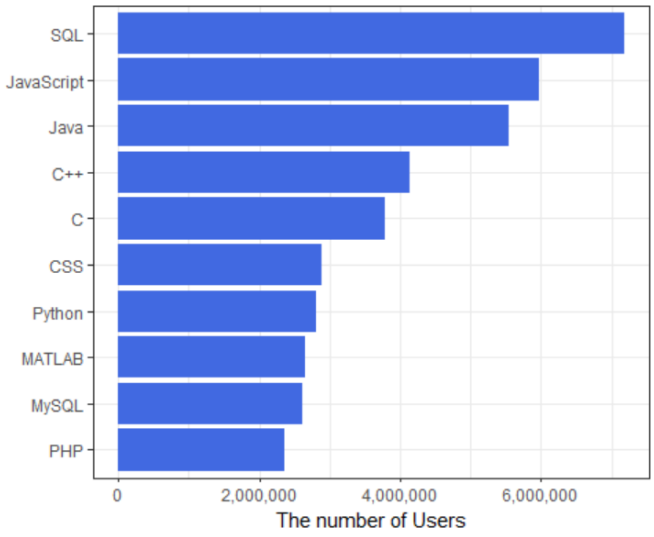

Figure 1: The top 10 languages used by users.

According to Figure 1, the top 10 languages include SQL, JavaScript, Java, C++, C, CSS, Python, MATLAB, MySQL, and PHP. Among these, SQL stands out as the most widely utilized language, with over 7,000,000 users, emphasizing its crucial role in managing and analysing structured data in various industries. JavaScript, with nearly 6,000,000 users, is highly prominent in web development, enabling the creation of dynamic and interactive web applications for both front-end and back-end purposes. General-purpose languages like Java, C++, and C continue to be important, finding applications in software development, system programming, and high-performance computing. The inclusion of CSS and MATLAB in the

top 10 highlights the significance of domain-specific languages tailored for specific tasks, such as web page styling and scientific computing, respectively. Python's popularity within the top 10 signifies its versatility in data analysis, machine learning, and artificial intelligence, owing to its extensive libraries and frameworks. MySQL reflects the enduring demand for relational databases and the need for robust data storage and retrieval systems. Lastly, PHP is a powerful, flexible, and adopted scripting language that empowers developers to create dynamic web applications, interact with databases, and automate tasks.

The importance of understanding the diverse landscape of languages and their respective applications can help guide decision-making regarding language selection for specific projects or career paths in the programming industry.

The top languages based on their ranking.

Figure 2: The top languages based on their ranking.

A word cloud graph shown in Figure 2 for languages with the most prominent words being Java, JavaScript, C++, PHP, Python, SQL, C, and Perl would suggest that these languages are widely used and relevant in the programming community. The ranking of languages can be influenced by several factors that impact their popularity such as industry demand, new technologies, community support, ease of use, job market. For instance, the rise of data science and machine learning has significantly increased the popularity of Python, while languages like Java and C++ continue to be widely used in enterprise software development.

It's important to note that the ranking of languages can change over time as technology evolves, new languages emerge, and industry demands shift. Therefore, it's crucial for developers to stay updated and adapt to changing trends and demands with the programming landscape.

The top 3 languages in the programming language and the query language that were required the most in job descriptions.

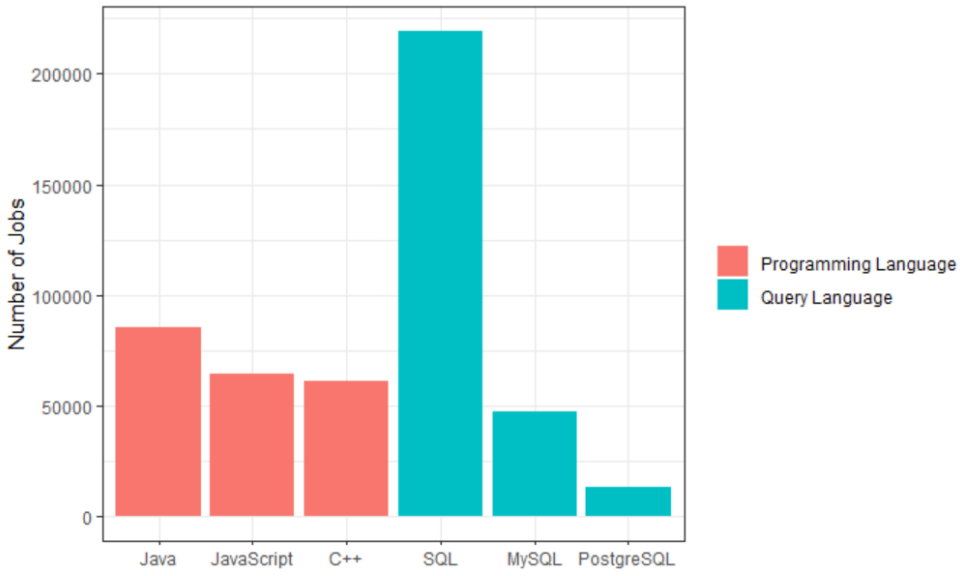

Figure 3: The top 3 languages were stated in job requirements.

The graph in Figure 3 illustrates the job demand in different programming and query languages. It focuses on three programming languages – Java, JavaScript, and C++, and three query languages – SQL, MySQL, and PostgreSQL. Java is the programming language that stands out among others due to its high demand, with an impressive number of over 85,000 job postings. Moving to JavaScript and C++, the graph shows that these two languages have a similar number of job requirements. Although not as dominant as Java, both languages still offer significant job opportunities, suggesting a strong demand for developers skilled in JavaScript and C++. When it comes to query languages, SQL takes the lead by a wide margin. With over 200,000 job requirements, SQL stands out as the preferred language for working with databases. In contrast, MySQL and PostgreSQL exhibit a noticeable difference in job demand. MySQL, with nearly 50,000 job requirements, enjoys a substantial presence in the job market, indicating its continued relevance in database management. On the other hand, PostgreSQL, although fewer in number with over 13,000 job requirements, still maintains a respectable position, suggesting a niche demand for PostgreSQL expertise.

The total number of languages deployed from 1952 to 2022.

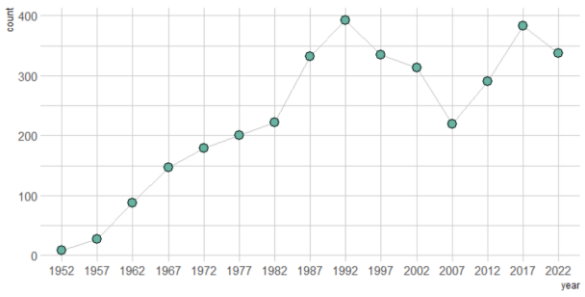

Figure 4: The number of languages were created from 1952 to 2022.

In Figure 4 above, the line chart depicts a gradual increase in the creation of languages over time, reflecting the growing needs and aspirations of programmers worldwide. In 1992, a remarkable peak is observed, with the number of languages reaching nearly 400. This period represents a significant milestone in language innovation. Following the peak in 1992, the number falls to around 200, indicating a notable decrease compared to the previous high. It might be considered that some existing languages such as Python, Java, JavaScript gain more popularity and adoption, which might have made it challenging for establishing new languages. However, starting from 2007, the line chart showcases a renewed upward trend, signifying a resurgence in language development.

The top languages referenced in the most published books across language types.

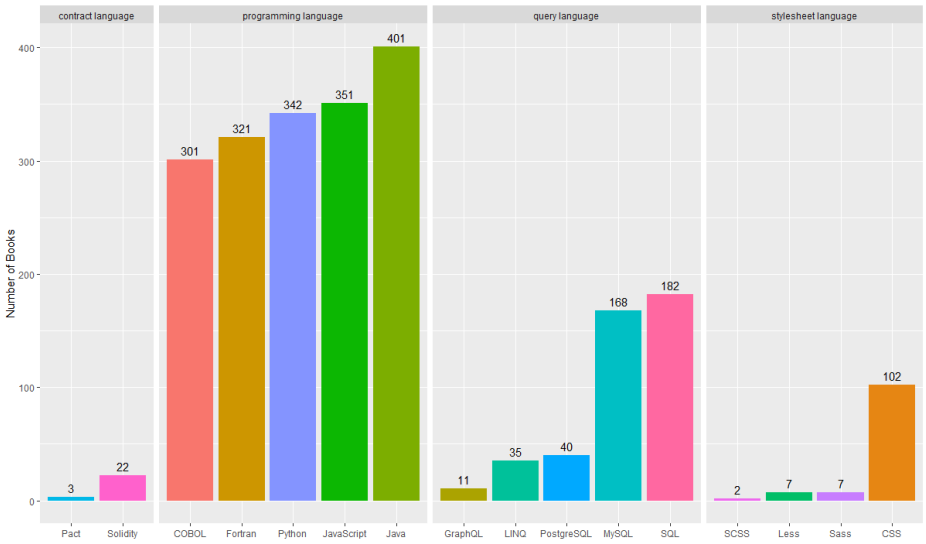

Figure 5: Number of Published Books per Language Group

In the realm of contract languages, there are only a few books available for Solidity and Pact. When it comes to programming languages, the highest number of published books is for Java, totalling 401. Following closely behind are JavaScript with 351 books, Python with 342 books, Fortran with 321 books, and COBOL with 301 books. In the domain of query languages, SQL emerges as the most prominent, with 182 published books. MySQL follows with 168 books, while PostgreSQL, LINQ, and GraphQL have relatively fewer published books. For style sheet languages, CSS stands out with 102 published books. On the other hand, Sass, Less, and SCSS have only a few books available.

The key takeaway from the provided information is that Java, JavaScript, and Python are the most popular programming languages in terms of the number of published books. SQL is the dominant query language, while CSS is the prominent style sheet language. This suggests that these languages and techniques have widespread adoption and significant resources available for learning and development. On the other hand, there are also emerging languages and technologies, such as Solidity and GraphQL, which may indicate potential growth and future relevance in development community.

The top 20 languages based on the number of answers.

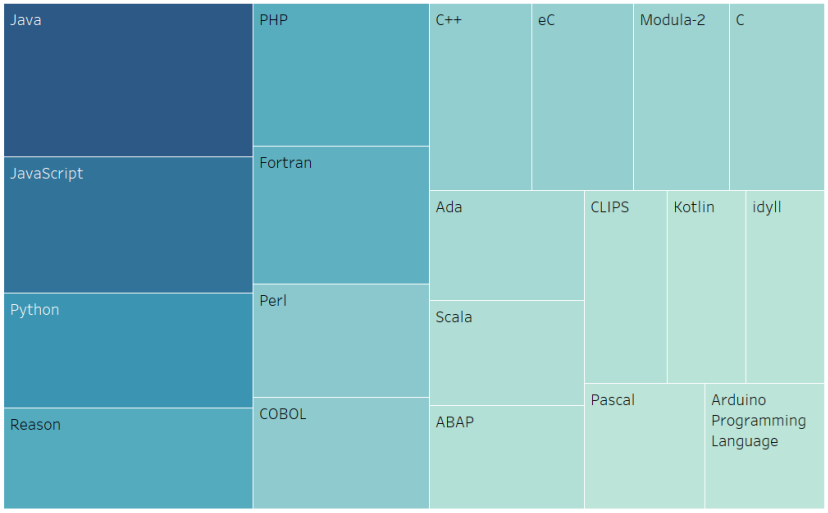

Figure 6: The top 20 languages based on the number of answers for each language.

According to Figure 6, this graph depicts the distribution of answers across various languages with respect to their support for problem-solving. It displays the top 20 languages, ranked based on the number of answers provided. At the forefront of the visualisation, Java stands tall as the leading language, boasting the highest number of answers. Following Java, JavaScript, Python, Reason, PHP, and Fortran emerge as prominent languages with a substantial count of answers, indicating their popularity and effectiveness in supporting problem-solving tasks. As the graph progresses downwards, the prominence of the languages decreases, signifying a relative lower number of answers. Positioned towards the bottom are Pascal and Arduino programming languages, which highlight their comparatively limited presence within the top 20.

Through this visual representation, viewers can easily grasp the distribution of answers across different languages and discern the varying levels of support they offer for problem-solving tasks.

The top countries created languages until 2022.

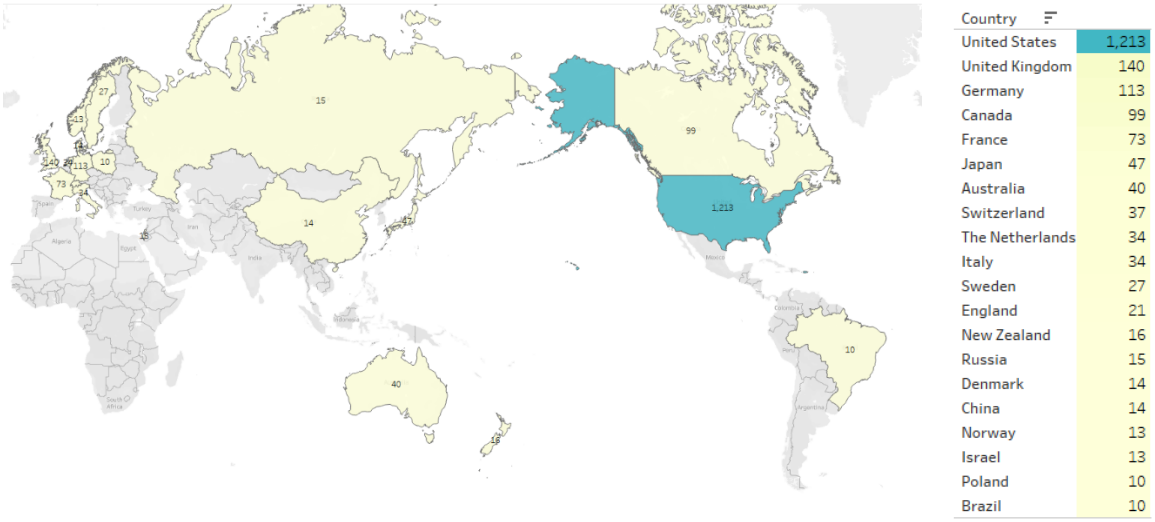

Figure 7: The top countries created the most languages until 2022.

In Figure 7 above, by highlighting the number of languages created by each country, particularly those with at least 10 creations, valuable insights are gained into the contributions and impact of these countries on the language landscape. The United States emerges as a significant player in this domain, with an astounding 1,213 languages created. This highlights the country's rich history of technological advancements and its role as a driving force behind language innovation. The United Kingdom has made substantial contributions, with 140 languages created. Germany's contribution of 113 languages further solidifies its position as a noteworthy participant in the language development. These countries demonstrate the country's influence and expertise in shaping the language landscape, particularly considering their smaller geographical size compared to the United States. Canada, with 99 languages, showcases its own share of significant contributions to the language ecosystem. Despite its relatively smaller population, the country has actively contributed to the development of languages, emphasising its dedication to technological innovation.

Hence, the country where a language was created can contribute to the growth of a vibrant developer community around that language. This includes the presence of user groups, conferences, educational institutions, open-source projects, and industry supports. These factors can influence the language's evolution, documentation, libraries, frameworks, and overall support. Figure 8 below clearly illustrates that the United States is the origin of numerous highly popular languages such as Java, JavaScript, C, SQL, and C++.

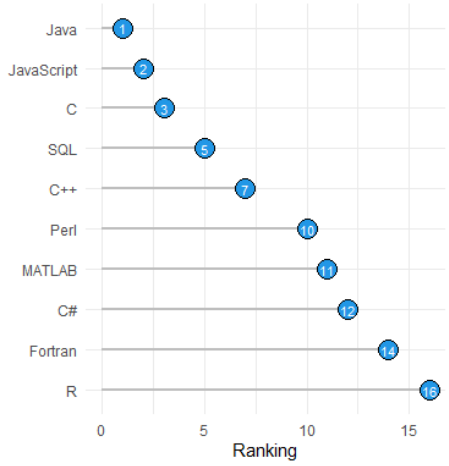

Figure 8: The top 10 ranking of languages created in the United States.

The languages have the top page views per day.

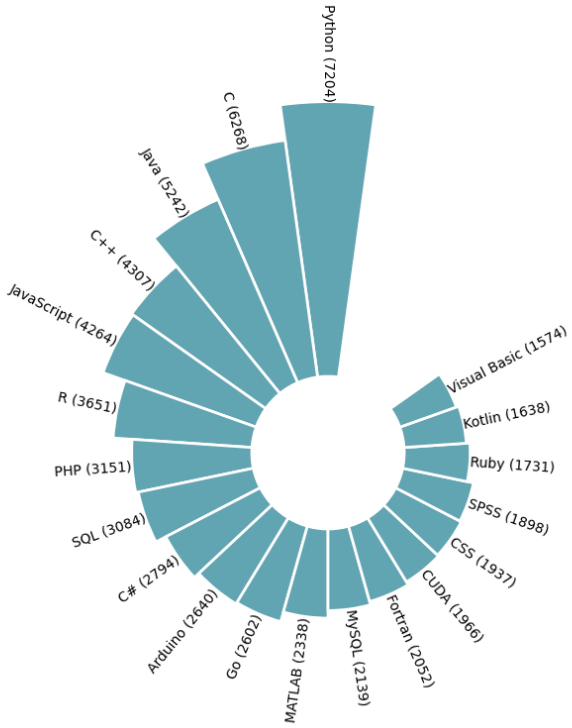

Figure 9: The top 20 languages have the most page views per day.

The Figure 9 above shows the popularity of 20 languages based on their daily page views. This analysis provides valuable insights into the language landscape and offers a glimpse into the preferences and interests of developers worldwide. Python emerges as the standout contender, amassing an impressive total of 7, 204 views. This indicates a remarkable level of interest and engagement with Python among developers. Python's success can be attributed to its versatility, ease to use, extensive libraries, and wide range of applications. Its popularity extends across diverse domains, including web development, data analysis, machine learning, and scientific computing. Notably, closely following Python are the languages C, Java, C++, and JavaScript, each maintaining a considerable viewership.

The presence of these languages in the upper echelons of page views highlights their enduring relevance and popularity in the development community. It signifies that developers continue to rely on them for a wide range of projects and applications.

Conclusion

As a result, the information from these visualisations offers valuable insights into the language landscape, including language popularity, job demand, problem-solving support, language creation trends, and country contributions. SQL, Java, JavaScript, C++, Python, and other languages have distinct roles and applications, reflecting the diversity of programming needs. Job demand is high for Java, JavaScript, C++, SQL, MySQL, and other languages, indicating their relevance in the industry. Language creation trends have shown growth and innovation, while the availability of resources and books highlight the popularity of certain languages. The distribution of problem-solving support and the contributions of countries like the United States and the United Kingdom further shape the language landscape. Overall, staying informed and adaptable to evolving trends is key in navigating the dynamic world of languages. Understanding the popularity of different languages can aid developers in making informed decisions regarding language selection, career prospects, and skill development. Moreover, it offers valuable information for organisations seeking to allocate resources effectively and make strategic decisions about technology adoption.

References

[1] J. Glenn Brookshear; Dennis Brylow; 2020; Computer Science: an overview.

[2] https://www.bls.gov/ooh/computer-and-information-technology/software-developers.htm

[3] https://www.kaggle.com/datasets/isaacwen/github-programming-languages-data

Appendix

The dataset is collected from the Kaggle website [3] based on the public domain database https://pldb.com. This dataset contains 4818 programming languages across 353 variables. The data source is large enough to enable the creation of a variety of visual representations that are both informative and engaging. According to the purpose of this study, sufficient variables are shown in the table below.

Variable	Description	Type
title	The official title of the language	Nominal
appeared	What year of the publicly released language	Interval
type	The category of the language	Nominal
languageRank	The rank amongst all languages in dataset	Ordinal

numberOfUsers	Number of users uses the language	Interval
numberOfJobs	Number of jobs requires the language	Interval
factCount	How many answers about the language	Interval
bookCount	How many published books for the language	Interval
wikipedia.dailyPageViews	How many page views per day for the language	Interval
country	What country the language was deployed in	Nominal

The data was analysed using R. The dataset was modified to contain variables in the table above by removing unnecessary columns. There are variety types of languages were inputted in the dataset. Based on this report, the dataset was filtered to select types at ‘pl’ (programming language), ‘queryLanguage’, ‘stylesheetLanguage’, ‘contractLanguage’.