BUSINESS BROADWAY
DATA SCIENCE | CUSTOMER ANALYTICS | MACHINE LEARNING

Search...

HOME   SERVICES ▾   RESOURCES ▾   ABOUT ▾   CONTACT   NEWSLETTER

For Data Professionals, Python Remains Top Programming Language while R Continues to Decline

by Bob Hayes on January 11, 2021 in Analytics, Data Science, Machine Learning

Data professionals of all stripes, including data scientists, machine learning engineers and others, use different types of tools in their jobs. A recent survey of over 20,000 data professionals by Kaggle revealed that Python, SQL and R continue to be the most popular programming languages. The most popular, by far, was Python (86% used). Additionally, 8 out of 10 data professionals recommended that aspiring data scientists learn Python first. R has seen a decline in usage over the past four years.

Kaggle conducted a worldwide survey in October 2020 of 20,036 data professionals (2020 Kaggle Machine Learning and Data Science Survey). Their survey included a variety of questions about data science, machine learning, education and more. Kaggle released the raw survey data and many of their members have analyzed the data (see link above). I will be exploring their survey data over the next couple of months. When I find something interesting, I'll be sure to post it here on my blog. Today's post is about the data science and machine learning programming languages data professionals used in 2020.
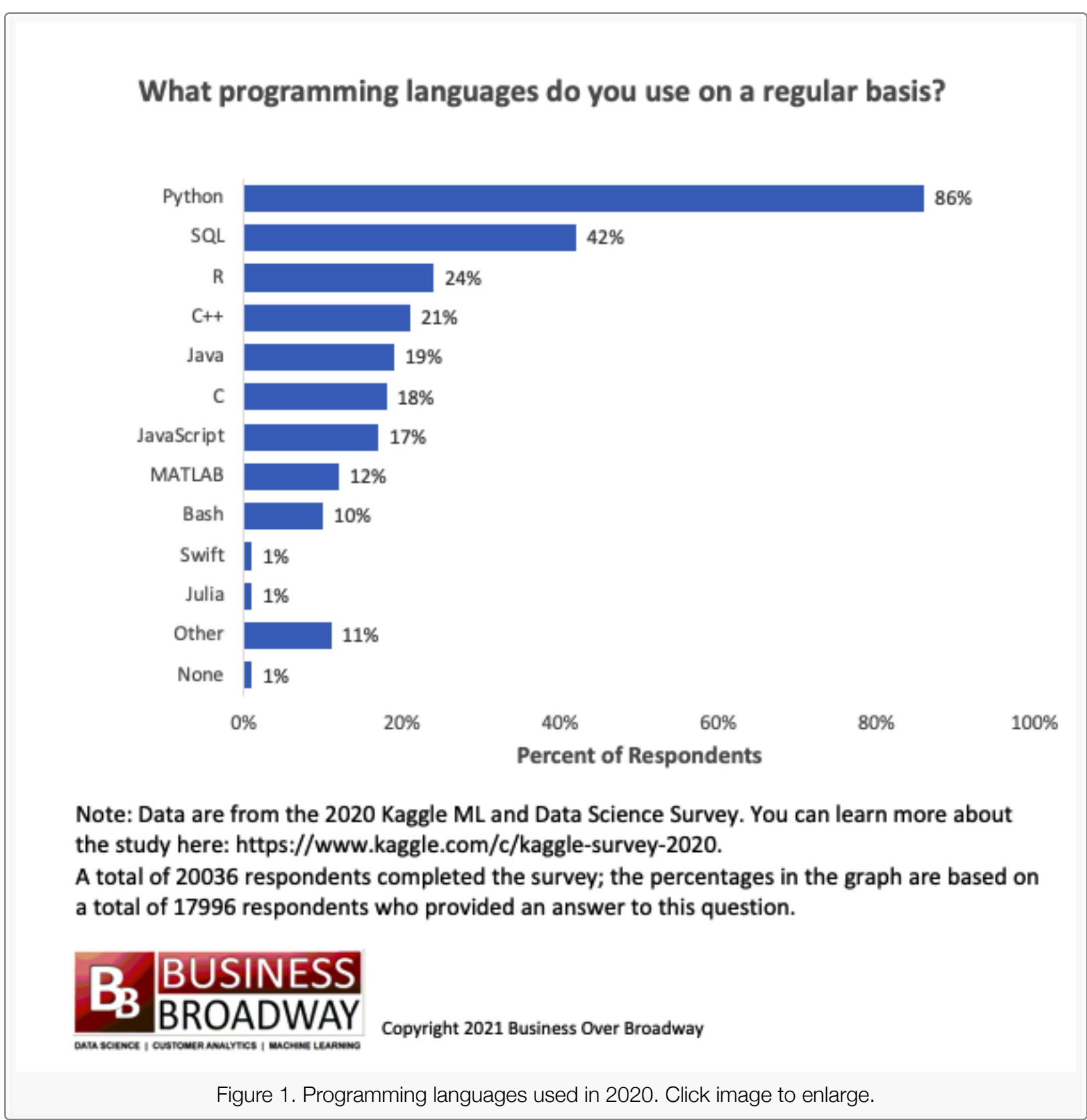

Figure 1. Programming languages used in 2020. Click image to enlarge.

Most Popular Programming Languages

The survey included a question, "What programming languages do you use on a regular basis? (Select all that apply)." On average, data professionals used 2 (median) languages in 2020. Data professionals who had the job titles of DBA/Database Engineer, Data Engineer and Software Engineer, however, used, on average 3 programming languages.

As seen in Figure 1, the top programming language in 2020 was Python (86% of respondents said they used this language), followed by SQL (42%) and R (24%). The top programming languages are rounded out by C/C++, Java, C, Javascript, MATLAB, Bash, Swift, and Julia.

Comparing program language usage from 2017, 2018 and 2019, we see that usage of Python has increased (60% used in 2017 and 83% used in 2018 and 87% in 2019) to 86%. SQL usage remains roughly the same over that same time period (42% used in 2017, 40% used in 2018, 44% used in 2019). However, R usage decreased over a 4-year period (usage was 46% in 2017, 36% in 2018 and 31% in 2019).

When we segment respondents based on their job title, we see that Python still reigns supreme across most of the job titles (see Figure 2). Top programming language used by Statisticians is R and the top programming language used by DBA/Database Engineers is SQL.

Of the data professionals who identified as a data scientist, 94% used Python, 56% used SQL and 37% used R.

Which Programming Language is Recommended Most?

The survey also asked respondents what programming language they would recommend an aspiring data scientist to learn first (see Figure 3). Results showed that 8 in 10 data professionals would recommend Python as the programming language aspiring data scientists to learn first. The remaining programming languages are recommended at a significantly lower rate (R recommended by 7% of respondents; SQL recommended by 5% of respondents.

When looking at different job titles, Python was, by far, the top recommended programming language. A minor exception was for Statisticians where 52% recommended Python and 34% recommended R.

Figure 2. Programming languages used by job title. Click image to enlarge.

Figure 3. Programming languages recommended most by data professionals for aspiring data scientists. Click image to enlarge.

Summary

The results of the Kaggle survey of about 20,000 data professionals clearly show the most popular programming languages for data professionals. Python, by far, continues to be the most popular programming language used by data professionals, followed by SQL and R. While Python maintains its popularity, R has seen a decline in use by data professionals over the past four years. Additionally, Python is the most recommended programming language for aspiring data writing paper professionals. Even though data professionals have access to many different programming languages, it appears that Python is the default programming language for data science and machine learning.

Subscribe

Subscribe to our e-mail newsletter to receive updates.

E-mail   SUBMIT

Related Posts:

- Data Professionals Prefer A Hybrid Approach to Tools and Technologies
- How Did the COVID-19 Pandemic Impact Collaboration among Data Scientists?
- Data Professionals are Dissatisfied with their Company's Use of Analytics
- Take the Data Science and Machine Learning Survey
- Gender Inequality Persists in Data Science and AI

🏷 Data science, machine learning, programming languages

Comments are closed.

About me

I am Business Over Broadway (B.O.B.). I like to solve problems through the application of the scientific method. I use data and analytics to help make decisions that are based on fact, not hyperbole. My interests are at the intersection of customer experience, data science and machine learning. To learn more about me and what I do, click here.

Tweets by @bobehayes

Bob E. Hayes Retweeted

Adam Parkhomenko ✔
@AdamParkhomenko

who made this

4h

Bob E. Hayes
@bobehayes

Was Your Information Swimming Naked During The Pandemic? @tdav @IBM

Why info couldn't be provided by IT:

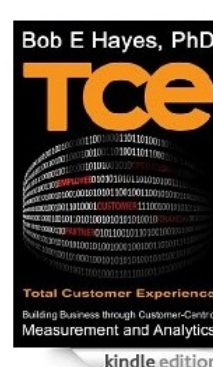1. The data could not be provided in needed timeframe 40%

Embed   View on Twitter

As Seen On

IBM   INTERNATIONAL YEAR OF STATISTICS   InformationWeek   customer THINK   KDD   Forbes   Inside CXM   cfpb

Happy Customers

EMC²   SENTRY   ORACLE   Fit4Way   VIRTUAL INSTRUMENTS   SOPHOS   Netsmart   illumina   CascadeMarketch   GENSTAR   Agilent Technologies   MOB4HERE
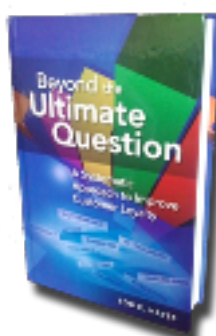

TCE: Total Customer Experience

Learn how to build your business around the customer using customer-centric measurement and analytics. Available in Kindle and paperback. Learn more about TCE: Total Customer Experience.

Beyond the Ultimate Question

Beyond the Ultimate Question: Book includes comprehensive review of Net Promoter® Score (NPS) and presents better measures of customer loyalty and best practices of customer feedback programs.

Measuring Customer Satisfaction

Measuring Customer Satisfaction and Loyalty (3rd Ed.): Book includes discussion of how to measure the customer experience, including customer satisfaction and customer loyalty.